DGK Ausschuss Geodäsie (DGK)
der Bayerischen Akademie der Wissenschaften

Reihe C | Dissertationen | Heft Nr. 951

Nico Adam

Improved SAR Coherence Magnitude Estimates in Scenarios with Low Coherence and Small Sample Size

München 2025

Verlag der Bayerischen Akademie der Wissenschaften, München

ISSN 0065-5325 ISBN 978-3-7696-5363-2

DGK Ausschuss Geodäsie (DGK)
der Bayerischen Akademie der Wissenschaften

Reihe C Dissertationen Heft Nr. 951

Improved SAR Coherence Magnitude Estimates in Scenarios with Low Coherence and Small Sample Size

Vollständiger Abdruck der von der Fakultät für Luft- und Raumfahrttechnik
der Universität der Bundeswehr München
zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

angenommenen Dissertation

von

Nico Adam

München 2025

Verlag der Bayerischen Akademie der Wissenschaften, München

ISSN 0065-5325 ISBN 978-3-7696-5363-2

Adresse des Ausschusses Geodäsie (DGK)
der Bayerischen Akademie der Wissenschaften:

Ausschuss Geodäsie (DGK) der Bayerischen Akademie der Wissenschaften
Alfons-Goppel-Straße 11 ● D – 80 539 München
Telefon +49 – 89 – 23 031 1113 ● Telefax +49 – 89 – 23 031 - 1283 / - 1100
e-mail post@dgk.badw.de ● http://www.dgk.badw.de

Vorsitzender: Univ.-Prof. Mag. Dr. habil. Thomas Pany
Gutachter: 1. Univ.-Prof. Dr.-Ing. habil. Michael Schmitt
2. Univ.-Prof. Dr.-Ing. Uwe Sörgel

Die Dissertation wurde am 19.09.2024 bei der Universität der Bundeswehr München eingereicht
und durch die Fakultät für Luft- und Raumfahrttechnik am 05.02.2025 angenommen.
Die mündliche Prüfung fand am 12.02.2025 statt.

Diese Dissertation ist auf dem Server des Ausschusses Geodäsie (DGK)
der Bayerischen Akademie der Wissenschaften, München unter <http://dgk.badw.de/>
sowie auf dem Server der Universität der Bundeswehr, München unter
<https://athene-forschung.unibw.de/doc/152204/152204.pdf> elektronisch publiziert

ISSN 0065-5325 ISBN 978-3-7696-5363-2

Executive Summary

Interferometric Synthetic Aperture Radar (SAR) (InSAR) is a well-established method for measuring the topography of the Earth and the displacements of its surface with millimeter accuracy. As this information is essential for infrastructure safety, there are numerous operational SAR missions, InSAR processing systems and InSAR-based monitoring services.

The SAR coherence magnitude is an essential parameter in InSAR. It is directly related to the signal-to-noise ratio and is therefore synonymous with it. It is also used to characterize InSAR systems, as the statistics of the interferometric measurements are parameterized by the coherence magnitude. And more recently, with SqueeSAR and CAESAR, it became the fundamental weighting for analyzing long time series from observations of distributed scatterers (DSs). The latter application, in particular, requires accurate coherence estimation to minimize error propagation and provide an accurate measurement of ground motion.

Estimators of the coherence magnitude, e.g., the sample estimator, are biased, and the smaller the coherence and the number of available samples, the more biased they are. The objective of this thesis is to develop new coherence magnitude estimators of jointly complex circular Gaussian (CCG) signals and to describe, characterize, and demonstrate the methods. In addition, Bayesian coherence priors are developed, which are applicable in everyday InSAR processing.

Bayesian methods are well-established in statistical inference and estimation, and allow to include prior information. However, there are currently no publications on methods for coherence magnitude estimation using this principle. Therefore, an empirical Bayesian estimation is developed. Another technique that has not yet been studied for coherence magnitude estimation is machine learning (ML). Two estimators are developed for this principle, and they are adapted to support prior information.

Using simulations, the estimators are characterized with respect to various sample sizes and the underlying true coherence by the corresponding bias, standard deviation, and root mean squared error (RMSE). Also, the respective performance is compared with the conventional sample estimator. Furthermore, the new methods are demonstrated on real Sentinel-1 data as a proof of concept.

In this thesis, the use of prior knowledge on the coherence is demonstrated for the first time, and all developed estimators support prior information. The more information is used and the stricter the prior, the more accurate the coherence estimate will be. The developed estimators offer two main advantages compared to the conventional sample

estimator. All improve the estimation of small coherences. And, they better estimate the coherence from small sample sizes. The empirical Bayesian estimator works advantageously up to 15 InSAR samples. The direct ML method is advantageous up to 30 samples, and the composite estimator was demonstrated to be advantageous for 200 InSAR samples. The performance and advantages are the reasons why the composite estimator is suitable and recommended for implementation in operational InSAR systems. It supports small and large sample sizes and has the best estimation performance compared to the other methods. An advantage worth emphasizing is its estimation performance even without prior. This makes the estimator universally applicable and comparable with the conventional sample estimator.

Kurzfassung

Die SAR-Interferometrie (InSAR) ist eine etablierte Methode zur Messung der Topografie der Erde und millimetergenauer Bewegungen ihrer Oberfläche. Da diese Informationen für die Sicherheit der Infrastruktur unerlässlich sind, gibt es zahlreiche operative SAR-Missionen, InSAR-Verarbeitungssysteme und InSAR-basierte Monitoring-Dienste für Bodenbewegungen.

Die Kohärenz ist ein wesentlicher Parameter bei InSAR. Sie steht in direktem Zusammenhang mit dem Signal-Rausch-Verhältnis und ist daher für dieses ein Synonym. Sie wird auch zur Charakterisierung von InSAR-Systemen verwendet, da die Statistik der interferometrischen Messungen durch die Kohärenz parametrisiert wird. Mit den neuen Prozessierungsverfahren SqueeSAR und CAESAR wurde es zur grundlegenden Gewichtung für die Analyse langer Zeitreihen aus Beobachtungen flächiger verteilter Streuer (auf engl. distributed scatterer). Insbesondere die letztgenannte Anwendung erfordert eine genaue Kohärenzschätzung, um eine Fehlerfortpflanzung zu vermeiden und eine präzise Messung der Bodenbewegung zu ermöglichen.

Alle Kohärenzschätzer, wie z. B. der Stichprobenschätzer, sind verzerrt, und zwar umso mehr, je kleiner die Kohärenz und die Anzahl der verfügbaren Stichprobenwerte ist. Das Ziel dieser Arbeit ist es, neue Kohärenzschätzer für *jointly complex circular Gaussian* (CCG) Signale zu entwickeln, die Methoden zu beschreiben, zu charakterisieren und zu demonstrieren. Darüber hinaus werden Bayes'sche Kohärenzprior entwickelt, die in der alltäglichen InSAR-Verarbeitung anwendbar sind.

Bayes'sche Methoden sind in der statistischen Inferenz und Schätzung fest etabliert und ermöglichen das Einbeziehen zusätzlicher Informationen. Allerdings gibt es bisher keine Veröffentlichungen über Methoden zur Kohärenzschätzung nach diesem Prinzip. Daher wird eine empirische Bayes'sche Schätzung entwickelt. Eine weitere Technik, die noch nicht für die Schätzung der Kohärenz untersucht wurde, ist das *Machine Learning* (ML). Für dieses Prinzip werden zwei Schätzer entwickelt, die an die Unterstützung von Vorabinformationen angepasst sind.

Durch Simulationen werden die Schätzer in Bezug auf verschiedene Stichprobengrößen und die zugrundeliegende tatsächliche Kohärenz durch den Bias, die Standardabweichung und den mittleren quadratischen Fehler (RMSE) charakterisiert. Zusätzlich wird die jeweilige Performanz mit dem herkömmlichen Stichprobenschätzer verglichen. Darüber hinaus werden die neuen Methoden anhand realer Sentinel-1 Daten als Machbarkeitsnachweis demonstriert.

In dieser Arbeit wird zum ersten Mal die Verwendung von Vorwissen über die Kohärenz demonstriert und alle entwickelten Schätzer unterstützen diesen Ansatz. Je mehr Information verwendet wird und je strenger der Prior ist, umso genauer wird die Kohärenzschätzung. Die entwickelten Schätzer bieten zwei wesentliche Vorteile im Vergleich zu den herkömmlichen Stichprobenschätzern. Alle verbessern die Schätzung von kleinen Kohärenzen. Und sie schätzen die Kohärenz bei kleinen Stichprobengrößen besser. Der empirische Bayes'sche Schätzer funktioniert vorteilhaft bis zu 15 InSAR-Stichprobenwerten. Die direkte ML-Methode ist bis zu 30 InSAR-Werten vorteilhaft, während der zusammengesetzte Schätzer sich bei 200 InSAR-Stichproben noch als vorteilhaft erwiesen hat. Die Leistung und die Vorteile sind die Gründe, warum der zusammengesetzte Schätzer für den Einsatz in operationellen InSAR-Systemen geeignet und empfohlen ist. Er unterstützt kleine und große Stichproben und hat im Vergleich zu den anderen Methoden die beste Schätzleistung. Ein hervorzuhebender Vorteil ist seine Schätzleistung auch ohne Prior. Dies macht den Schätzer universell einsetzbar und vergleichbar mit dem herkömmlichen Stichprobenschätzer.

Contents

List of Symbols

Latin Symbols

$\mathbf{1}$	vector of ones
$\boldsymbol{A}$	decomposition of the square, positive definite and Hermitian covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}^H$
$\boldsymbol{a}$	vector of square roots of the expected intensities I_k of each channel k, $\boldsymbol{a} = \left(\sqrt{I_1} \quad \sqrt{I_2} \quad \dots \quad \sqrt{I_{N_{\mathrm{SLC}}}}\right)^T$
a	amplitude without relation to a specific sample, e.g., an argument of a PDF
$\hat{\boldsymbol{a}}$	estimate of $\boldsymbol{a}$, e.g., from square root of diagonal of estimated covariance matrix
a_1	primary scene amplitude of a sample used, e.g., as argument of a PDF
a_2	secondary scene amplitude of a sample used, e.g., as argument of a PDF
$a_{k,i}$	SLC scene pixel amplitude $a_{k,i} = \lvert x_{k,i}\rvert$ with scene index k and pixel index i
B_{eff}	effective, i.e., perpendicular, baseline between primary and secondary sensor
D	data sample
$\mathbb{I}$	identity matrix
I_1	expected (true) intensity of primary channel in $\mathrm{m}^2/\mathrm{m}^2$
I_2	expected (true) intensity in $\mathrm{m}^2/\mathrm{m}^2$ of secondary channel
I_k	expected (true) intensity in $\mathrm{m}^2/\mathrm{m}^2$ of channel with index k
$\hat{I}_k$	estimated intensity in $\mathrm{m}^2/\mathrm{m}^2$ of channel with index k
i	pixel index within the SHP
j	imaginary number $\mathrm{j} = \sqrt{-1}$
k	scene index, e.g., $k = 1$ for the primary and $k = 2$ for the secondary channel
k_1	index of primary SLC scene and in coherency matrix first index
k_2	index of secondary SLC scene and in coherency matrix second index
k_z	vertical wavenumber or, more precisely, phase to height conversion factor, in m/rad resulting from observation geometry
L	number of independent coherence estimates
M	number of random samples $\boldsymbol{X}^{**}_{r,m}$ in Double Bootstrapping
m	normalization constant of non-normalized posterior function
$\underline{m}_1$	first log-moment
N	sample size, i.e., number of i.i.d. samples in a SHP of a single scene
N_{pixel}	number of homogenous SAR scene samples of a single scene affected by spectral weighting and oversampling
N_{SLC}	number of radar scenes which are SLC data
$P_{N,S}$	number of independent subsamples each of size S from N total samples

R	number of random samples $\boldsymbol{X}_r^*$ in Double Bootstrapping
R_0	range distance from sensor to ground
$R_{k,i}$	range distance from sensor to the resolution cell phase center in km including a change of distance in mm caused by ground motion
r	oversampling of the data
r_{az}	oversampling of the data in azimuth direction
r_{rg}	oversampling of the data in range direction
$\boldsymbol{S}$	complex matrix with simulated interferometric data pair $\boldsymbol{S} \in \mathbb{C}^{2\times N}$
S	size of a subsample
SNR_1	signal-to-noise ratio of primary scene
SNR_2	signal-to-noise ratio of secondary scene
SNR_k	signal-to-noise ratio of channel with index k
$\boldsymbol{T}$	coherency matrix
T	total number of samples $T = NL$
$\hat{\boldsymbol{T}}$	estimated coherency matrix
$\hat{\boldsymbol{T}}_{\mathrm{m}}$	modified coherency matrix, i.e., estimated phases are kept, however coherences are changed
$\hat{\boldsymbol{T}}_{\mathrm{s}}$	coherency matrix estimated from sample estimator
t	elapsed time in d
$\boldsymbol{U}$	left eigenvectors from SVD
$\boldsymbol{u}_i$	eigenvector with index i related to eigenvalue λ_i from eigendecomposition
$\boldsymbol{V}$	right eigenvectors from SVD
$\boldsymbol{W}$	diagonal matrix with singular values from SVD
$\boldsymbol{X}$	zero mean complex N_{SLC}-variate Gaussian random variable representing SHP according to Fig. 1
$\boldsymbol{X}_1$	CCG process of the primary channel
$\boldsymbol{X}_2$	CCG process of the secondary channel
$\boldsymbol{X}_r^*$	random interferometric sample from $\boldsymbol{X}$ in Double Bootstrapping
$\boldsymbol{X}_{r,m}^{**}$	random interferometric sample from $\boldsymbol{X}_r^*$ in Double Bootstrapping
$\boldsymbol{x}$	column vector at any position i, i.e., along the scenes, $\boldsymbol{x} = \boldsymbol{X}_{*,i} = \begin{pmatrix} x_{1,i} & \dots & x_{k,i} & \dots & x_{N_{\mathrm{SLC}},i} \end{pmatrix}^T$
$\boldsymbol{x}_1$	complex samples of primary scene, $\boldsymbol{x}_1 = \boldsymbol{X}_{1,*} = \begin{pmatrix} x_{1,1} & \dots & x_{1,i} & \dots & x_{1,N} \end{pmatrix}^T$
$\boldsymbol{x}_2$	complex samples of secondary scene, $\boldsymbol{x}_2 = \boldsymbol{X}_{2,*} = \begin{pmatrix} x_{2,1} & \dots & x_{2,i} & \dots & x_{2,N} \end{pmatrix}^T$
$\boldsymbol{x}_k$	complex samples of scene with index k, $\boldsymbol{x}_k = \boldsymbol{X}_{k,*} = \begin{pmatrix} x_{k,1} & \dots & x_{k,i} & \dots & x_{k,N} \end{pmatrix}^T$
$x_{k,i}$	SLC scene pixel with scene index k and pixel index i within the SHP, e.g., $k = 1$ for the primary and $k = 2$ for the secondary scene and $i = 1 \dots N$. The symbols for the complex pixel representation are $x_{k,i} = a_{k,i}\, \mathrm{e}^{\mathrm{j}\,\phi_{k,i}}$
$\boldsymbol{x}_{\mathrm{PS}}$	representation of the SHP $\boldsymbol{X}$ by a PS like data vector with unit amplitude for all scenes, $\boldsymbol{x}_{\mathrm{PS}} = \begin{pmatrix} \mathrm{e}^{\mathrm{j}\phi_1} & \dots & \mathrm{e}^{\mathrm{j}\phi_k} & \dots & \mathrm{e}^{\mathrm{j}\phi_{N_{\mathrm{SLC}}}} \end{pmatrix}^T$

$\hat{\boldsymbol{x}}_{\mathrm{PS}}$	estimate of a representation of the SHP $\boldsymbol{X}$ by a PS like data vector with unit amplitude for all scenes
$\boldsymbol{y}_{\mathrm{PS}}$	representation of the SHP $\boldsymbol{X}$ by a PS like data vector, $\boldsymbol{y}_{\mathrm{PS}} = \left(a_1 \mathrm{e}^{\mathrm{j}\phi_1} \quad \dots \quad a_k \mathrm{e}^{\mathrm{j}\phi_k} \quad \dots \quad a_{N_{\mathrm{SLC}}} \mathrm{e}^{\mathrm{j}\phi_{N_{\mathrm{SLC}}}}\right)^T$
$\boldsymbol{Z}$	complex matrix of independent CCG random variables each channel with power 1
z_i	complex interferogram sample $z_i = x_{1,i}\, x_{2,i}^*$ with pixel index i within the SHP generated from a primary and secondary scene
$z_{k,i}$	complex matrix element of $\boldsymbol{Z}$

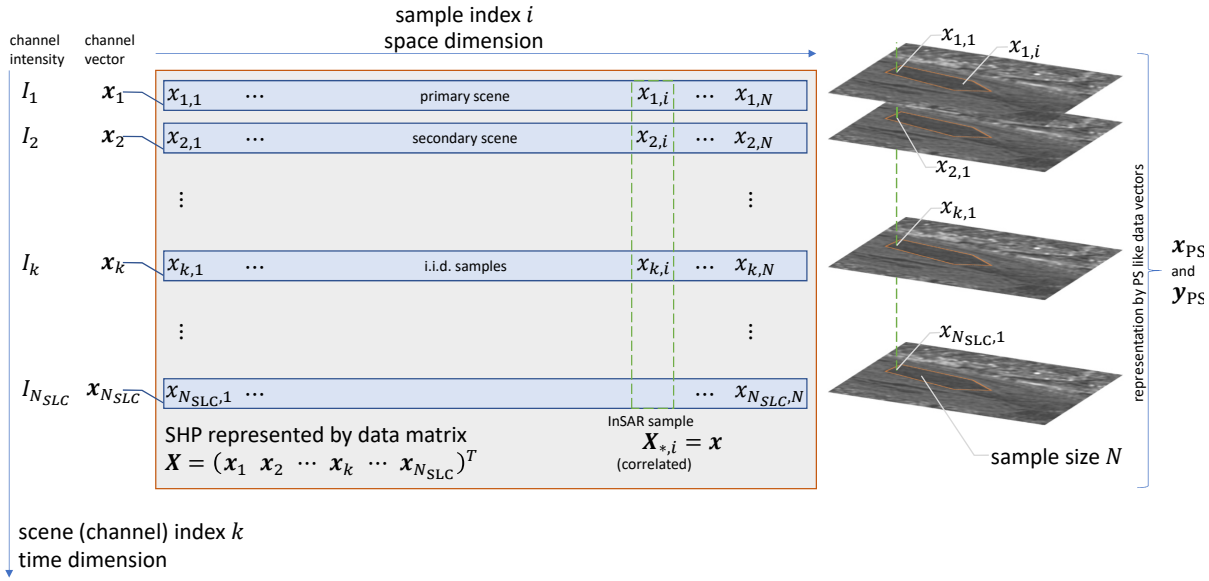


Figure 1: Data representation and respective symbols.

Greek Symbols

β_k^0	radar brightness (Beta Naught) of channel with index k
γ	(true) absolute value of the complex coherence
γ_∞	long-term coherence
$\gamma_*^{(\mathrm{bias})}$	bias of coherence estimator with name $*$
$\gamma_*^{(\sigma)}$	standard deviation of coherence estimator with name $*$
$\gamma_*^{(\mathrm{RMSE})}$	RMSE of coherence estimator with name $*$
$\gamma_{k_1,k_2}^{(\mathrm{simu})}$	coherence magnitude used in a simulation for scenes with index k_1 and k_2
$\gamma_{\max}$	maximum coherence parameter of prior
γ_{simu}	coherence magnitude value used in simulation
$\hat{\gamma}$	estimated coherence magnitude

$\hat{\gamma}_{\mathrm{CCM}}^{(L)}$	coherence magnitude from a complex coherence map
$\hat{\gamma}_{\mathrm{CMM}}^{(L)}$	coherence magnitude from a coherence magnitude map
$\hat{\gamma}_{\mathrm{DB}}$	estimated coherence magnitude from Double Bootstrapping
$\hat{\gamma}_{\mathrm{EAP}}$	estimated coherence magnitude from expected a posteriori (EAP) estimator
$\hat{\gamma}_{\mathrm{LCMM}}^{(L)}$	coherence magnitude from a logarithmic coherence magnitude map
$\hat{\gamma}_{\mathrm{MAP}}$	estimated coherence magnitude from maximum a posteriori (MAP) estimator
$\hat{\gamma}_{\mathrm{MEDAP}}$	estimated coherence magnitude from median a posteriori (MEDAP) estimator
$\hat{\gamma}_r^*$	coherence magnitude sample estimate from random sample $\boldsymbol{X}_r^*$
$\hat{\gamma}_{r,m}^{**}$	coherence magnitude sample estimate from random sample $\boldsymbol{X}_{r,m}^{**}$
$\hat{\gamma}_{\mathrm{s}}$	estimated coherence magnitude by sample estimator
$\hat{\gamma}_{\mathrm{s},i}$	value at index i from set of coherence magnitudes from sample estimator
Δ	mutual shift between primary and secondary scene
$\hat{\Delta}$	estimate of mutual shift Δ from complex cross correlation
ϵ	inherent random error term (irreducible error)
θ_0	local incidence angle between LOS and surface normal
λ	radar wavelength, e.g., TSX 31.1 mm, Sentinel-1 55.5 mm, ERS 56.6 mm
λ_i	eigenvalue with index i related to eigenvector $\boldsymbol{u}_i$ from eigendecomposition
$\rho_{\boldsymbol{X}_1,\boldsymbol{X}_2}$	linear correlation coefficient between processes $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$
Σ	covariance matrix
σ	standard deviation of real and imaginary component of a SLC scene without relation to a specific sample used, e.g., as argument of a PDF
$\sigma_{\hat{\Delta}}$	standard deviation of mutual shift estimate from complex cross correlation
$\sigma_{\hat{\Delta}}^{(\mathrm{bias})}$	bias of $\sigma_{\hat{\Delta}}$-estimate based on a sample coherence estimate
$\sigma_{\hat{h}}$	standard deviation of estimated topography height
σ_{im}	standard deviation of the imaginary component of a SLC scene
σ_k	standard deviation of real and imaginary component of a SLC scene $\sigma_k = \sigma_{\mathrm{re}} = \sigma_{\mathrm{im}}$ which is related to the channel intensity $I_k = 2\sigma_k^2$
σ_k^0	backscatter coefficient (Sigma Naught) of channel with index k
σ_{re}	standard deviation of the real component of a SLC scene
$\sigma_{\varphi_{1,2}}$	standard deviation of the interferometric phase of a single InSAR data pair
$\sigma_{\hat{\varphi}_{\mathrm{s}}}$	standard deviation of the interferometric multilook phase
τ	characteristic decorrelation time in d
$\phi_{k,i}$	SLC sample phase $\phi_{k,i} = \arg(x_{k,i})$ in rad (scene index k and pixel index i)
$\phi_{k,i}^{(\mathrm{noise})}$	noise phase component of $\phi_{k,i}$, e.g., from thermal noise
$\phi_{k,i}^{(\mathrm{obj})}$	random object phase component of $\phi_{k,i}$
$\phi_{k,i}^{(\mathrm{R})}$	signal phase component of $\phi_{k,i}$ which is related to the range distance $R_{k,i}$; forming the interferogram, this component results in InSAR phase measurements $\varphi_i^{(\mathrm{h})} + \varphi_i^{(\mathrm{defo})} + \varphi_i^{(\mathrm{atmo})}$
φ	(true) phase of the complex coherence
$\varphi_{1,2}$	interferometric phase difference in rad from a primary and secondary scene
φ_i	interferogram pixel phase in rad with pixel index i within the SHP generated from a primary and secondary scene
$\varphi_i^{(\mathrm{atmo})}$	interferometric phase component resulting from atmospheric effects in the primary and secondary scene

$\varphi_i^{(\mathrm{defo})}$	interferometric phase component resulting from deformation
$\varphi_i^{(\mathrm{h})}$	interferometric phase component resulting from topography
$\varphi_i^{(\mathrm{noise})}$	interferometric phase noise
$\hat{\varphi}_{k_1,k_2}$	estimate of interferometric phase from scenes with index k_1 and k_2
$\hat{\varphi}_{\mathrm{s}}$	sample estimate of interferometric phase in rad $\hat{\varphi}_{\mathrm{s}} = \arg\left(\sum_{i=1}^{N} x_{1,i}\, x_{2,i}^{*}\right)$

Functions and Operators

$\circ$	Hadamard product
$\oslash$	Hadamard division
$\lvert . \rvert$	absolute value of a complex or real variable
$\lfloor . \rfloor$	*floor* function
$\sim$	$x \sim \mathrm{p}(.)$ means x is distributed according to or sampled from $\mathrm{p}(.)$
${}_2\mathrm{F}_1(.)$	Gaussian hypergeometric function
${}_3\mathrm{F}_2(.)$	generalized hypergeometric function
$\Gamma(.)$	gamma function
$\mathrm{Cov}(.,.)$	covariance of arguments which are random processes or random samples
$\mathrm{E}\{.\}$	expected value of argument which is a random process or random sample
$\mathrm{Li}_2(.)$	dilogarithm function
$\mathfrak{Im}(.)$	imaginary part of argument
$\mathcal{N}(\mu, \sigma)$	normal distribution with mean μ and standard deviation σ, i.e., $\mathcal{N}(\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$
$\mathrm{P}(.)$	CDF
$\mathrm{p}(.)$	PDF
$\mathfrak{Re}(.)$	real part of argument
$\mathrm{U}(a, b)$	continuous uniform distribution with range $a \ldots b$
$.^T$	transpose of a vector or matrix
$.^H$	conjugate transpose (Hermitian transpose) of a vector or matrix

SI Units

d	unit for time in days
km	unit for length in kilometers
mm	unit for length in millimeters
rad	unit for angle in radian
yr	unit for time in years

1 Introduction

1.1 Scientific Motivation and Relevance of the Topic

Interferometric SAR (InSAR) is well-established for surveying the topography of the Earth (Hensley et al. 2000, Eineder et al. 2001, Zink et al. 2017) and its changes, i.e., ground deformations (Ferretti et al. 2001, Berardino et al. 2002, Adam et al. 2003, Hooper et al. 2004, Guarnieri & Tebaldini 2008, Ferretti et al. 2011). The reasons for this are, on the one hand, cost-effectiveness due to remote sensing surveying and automatic data processing and, on the other hand, the characteristics of the data. The generated digital elevation model (DEM) is homogeneous and consistent as it has a uniform pixel grid, is acquired over a large area during a short mission time, and the observation is independent of cloud cover in contrast to optical methods.

In recent years, InSAR has developed rapidly and now allows continuous monitoring of subtle deformations of the Earth's surface with millimeter accuracy. The basis is the persistent scatterer interferometry (PSI), which was invented by Ferretti et al. (2001). As an example, Fig. 1.1 visualizes the ground deformation in the area nearby Cologne

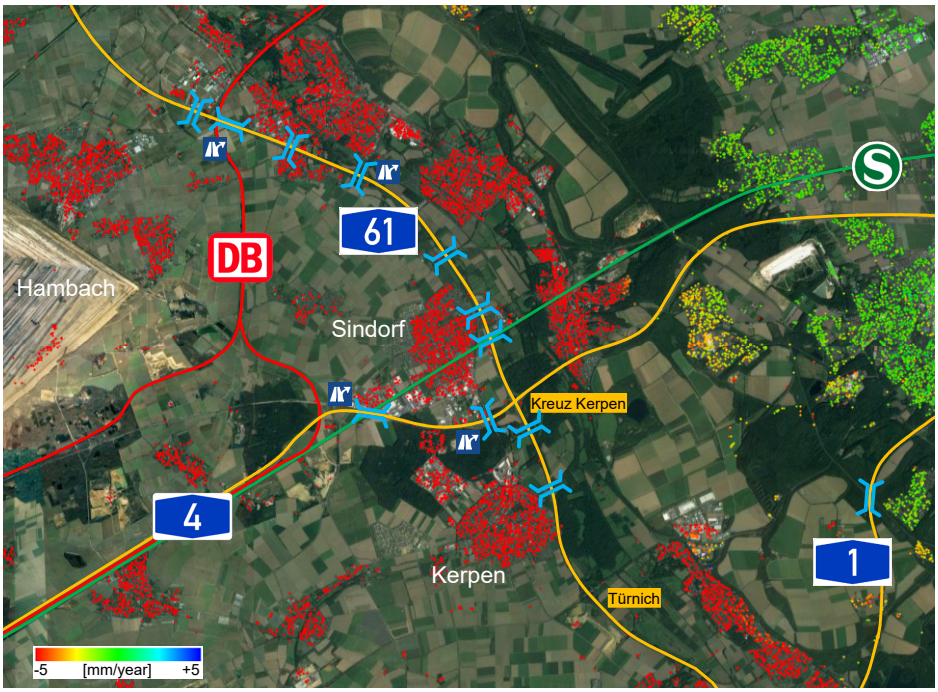


Figure 1.1: Ground motion monitoring example of infrastructure from ERS data.

(Germany), with red indicating strong subsidence of more than 30 mm/yr and green indicating stable areas. The spaceborne monitoring of ground motion with millimeter range is important because such information is crucial for the safety of infrastructure and buildings as well as for the early detection of geohazards. The example above illustrates the amount of omnipresent infrastructure worth protecting in a small area of ca. 20 km $\times$ 15 km with railway lines, highways, bridges and buildings.

There is an increasing number of wide area operational services such as the European Ground Motion Service (EGMS) (Frei 2017, Crosetto et al. 2020, Costantini et al. 2022), and the Ground Motion Service Germany (BBD) (Kalia 2017, Kalia et al. 2018, Lege et al. 2019) that make the deformation maps freely available and, thus, widely visible (Even et al. 2024). For their production, the coherence magnitude is an essential estimate. As pointed out by Zebker & Villasenor (1992) as well as Just & Bamler (1994), the coherence magnitude is a proxy for the signal-to-noise ratio (SNR). This is the reason why it is the crucial weighting in all estimation methods based on distributed scatterers. Furthermore, Bamler & Hartl (1998) describe the coherence to be a useful concept for InSAR system description and for interferogram quality assessment. Due to the large amount of data and the significance, there is an actual need to estimate this parameter as accurately and computationally effectively as possible. Technically, the task is to estimate the population parameter coherence magnitude from a sample of size N. However, the challenges are the bias and variance of the estimate, which are large for small coherences and small sample sizes, and the high computational cost of using more precise methods. Up to now, little is published on the Bayesian estimation, i.e., include prior information, on using machine learning (ML) and on composite inference methods for the estimation of the Synthetic Aperture Radar (SAR) coherence magnitude.

1.2 Problem Statement and Objectives

The objective of this thesis is to develop new SAR coherence magnitude estimators and to describe the principle and respective methods. Using simulations, they are characterized with respect to various sample sizes and the underlying true coherence by the corresponding bias, standard deviation, and root mean squared error (RMSE). Also, the respective performance is compared with the conventional sample estimator. Furthermore, the new methods are demonstrated on real data as a proof of concept.

Bayesian methods are well-established in statistical inference and estimation and allow to include prior information. However, there are currently no publications on methods for SAR coherence estimation using this principle. In this thesis, the use of prior knowledge on the coherence is demonstrated for the first time with the newly developed empirical Bayesian coherence magnitude estimation.

Another objective of this thesis is to improve the empirical Bayesian coherence estimation in terms of accuracy and computational cost. Up to now, the coherence magnitude estimation based on ML has not been studied and published. In this thesis, the ML framework is applied to estimate the SAR coherence magnitude. Two types of ML algorithms

are developed. The first method implements the fundamental principle and already improves the estimation performance. The second method is more complex and uses the first method as a subsystem, but improves the estimate even further. The principle of prior is established also for the ML methods. This means, this thesis solves the problem how to integrate the prior information into the ML approach.

In this thesis, coherence priors are developed which are applicable in everyday InSAR processing. Uninformative and informative priors are accomplished and for the latter, two typical InSAR scenarios are being considered, i.e., a strict prior and a less strict prior. In order to support a practical application, the techniques are checked for different use cases whether they are suitable for operational systems in terms of estimation performance and availability of priors.

1.3 Solution Approach and Methodology

In this thesis, the coherence estimation is systematically improved using various techniques. Figure 1.2 visualizes the developed methods and highlights them with reddish boxes. All three methods are based on different principles. Firstly, Bayesian inference is an established method in parameter estimation. However, it has not yet been used for coherence estimation. Furthermore, Bayesian estimation is understood as a regularization of the maximum likelihood estimation, which is currently the state of the art in coherence estimation. Second, ML is an area of research that is currently receiving a lot of attention, especially for classification and regression analysis for predictive modelling. Methods of the latter are adapted and used for coherence estimation and the inclusion of prior information. Thirdly, composite estimation is a further improvement of coherence estimation. In this method, the ML framework learns the statistics of partial estimates from subsamples. It utilizes the performance and accuracy of the second ML algorithm. An important contribution of this thesis is that the concept of prior information has been introduced into the SAR coherence estimation and has been systematically implemented in all techniques. The more information is used and the stricter the general prior, the more accurate the estimate will be.

1.4 Thesis Outline

The remainder of this thesis is organized as follows. Section 2 introduces InSAR with some background information and fundamentals, e.g., InSAR theory and data statistics, coherence applications, and the state of the art in estimating the coherence magnitude. Section 3 presents the methods of all estimators developed. Section 4 provides the experimental results, including a Sentinel-1 test case for each method. Finally, Section 5 presents the conclusion of this thesis and an outlook on possibilities for further development of the research conducted in this thesis.

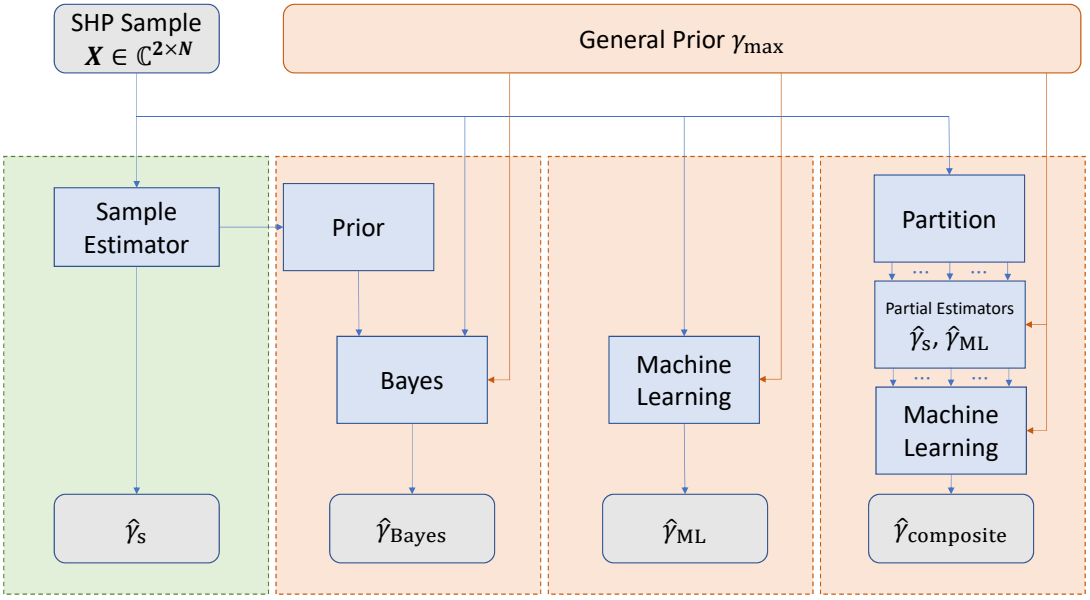


Figure 1.2: Overview about the developed coherence estimation methods.

2 SAR Interferometry

2.1 InSAR Background and Fundamentals

InSAR is a modern radar remote sensing application. It is based on SAR as coherent imaging method. Bamler & Hartl (1998) provide in detail the technology and the signal theoretical aspects of InSAR as well its origins (Richman 1971, Graham 1974, Goldstein et al. 1985). Since then, InSAR has developed into an active research field with numerous operational spaceborne missions e.g., Seasat, ERS-1/2, JERS-1, SRTM, Envisat/ASAR, RadarSAT-1/2, TerraSAR-X, TanDEM-X and the fleets of Radarsat, Cosmo-SkyMed, Hongtu-1 and of Sentinel-1. The step from research to operational commercial missions has now been completed, e.g., by Capella's, ICEYE's and Umbra's SAR satellite constellations.

First InSAR processing results were published by Zebker & Goldstein (1986), Gabriel & Goldstein (1988) and Prati et al. (1989). With across-track InSAR, the phase information contained in the radar echoes is used to calculate the topography of the Earth or displacements of the Earth's surface from at least two acquisitions. For topography mapping, the radars of the acquisitions need to be spatially separated perpendicular to the line of sight (LOS). For deformation monitoring, the acquisitions have to be separated in time and the technique is named differential InSAR (D-InSAR). This method was first demonstrated in practice by Massonnet et al. (1993) and Goldstein et al. (1993). PSI invented by Ferretti et al. (2001) is an operational spaceborne SAR technique to monitor subtle motion of the Earth's surface. It extends the InSAR by a time series analysis and allows mitigation of atmospheric effects and as a result millimeter monitoring precision. Variants of multi-temporal techniques are developed by Berardino et al. (2002), Colesanti et al. (2003), Adam et al. (2003), Hooper et al. (2004), Kampes (2005), Ferretti et al. (2011), Fornaro et al. (2015) and Ansari et al. (2017).

Another extension is SAR tomography (Reigber & Moreira 2000, Fornaro et al. 2003, Lombardini 2003, Zhu & Bamler 2010). It reconstructs the reflectivity along elevation and separates scatterers within a resolution cell.

The subsequent subsections explain the InSAR principle and the signals as well as their statistics, which are typically expressions parameterized by coherence magnitude. This is the reason why Bamler & Hartl (1998) describe the coherence to be a useful concept for system description and for interferogram quality assessment. In the sections below, some statistics and an experiment are presented to demonstrate the importance of this parameter and applications of its estimate.

2.1.1 Data Acquisition Principle

The SAR illuminates an area on the ground. In doing this, the reflected echoes are recorded and processed coherently. The result is a radar image that is composed of complex numbers. Figure 2.1 shows on the right an example radar acquisition with the amplitude to explain the two types of backscatters that can be observed. Distinguishing these scattering types is important because they are characterized differently.

Point scatterer is a dominant scatterer with a high reflectivity and a fix phase center which is only slightly affected by temporal and geometrical decorrelation. As a consequence, such a scatterer has a high phase stability over long time spans and different look angles (effective baselines). Ferretti et al. (1999) introduced point scatterers (PSs) into a systematic interferometric processing. The used scatterers are characterized by di- or trihedral reflection or simple deterministic antenna like single-bounce scattering. Typically, man-made structures made of metal, boulders, and outcrops can all generate good PSs (Ferretti et al. 2011). PSs are characterized by the signal-to-clutter ratio (SCR) (Freeman 1992) which is related to and estimated by the dispersion index (Ferretti et al. 2001).

Distributed scatterer is an areal scatterer observed from vegetated areas, fields and forests. If the scattering surface is rough with respect to the radar wavelength λ and the resolution cell includes many scatterers, i.e., it is large compared to the single scatterers, the amplitude and phase of all elementary scatterers are statistically independent and all

Figure 2.1: SAR scene amplitude (right image), point scatterer example (upper left image) and distributed scatterer example (lower left image).

phases are equally likely in the interval $[-\pi, \pi)$. It is characterized by a homogeneous backscattering intensity. Such data are modeled by a stationary complex, circular Gaussian (CCG) process as stated by Goodman (1976) and Just & Bamler (1994) as well as Rocca (2007). In practice, distributed scatterers (DSs) are not coherent over long temporal intervals. Nevertheless, they can be exploited for motion estimation using conventional D-InSAR techniques Rocca (2007) and advanced time series methods (Berardino et al. 2002, Ferretti et al. 2011, Ansari et al. 2017).

2.1.2 InSAR Theory

The variable $x_{k,i} = a_{k,i} \exp(\mathrm{j}\phi_{k,i})$ denotes the single look complex (SLC) SAR scene pixel with acquisition index k, e.g., $k = 1$ for the primary and $k = 2$ for the secondary scene, more technically named as channel. i is the pixel index within a statistically homogeneous area with N independent and identically distributed (i.i.d.) samples. In this thesis, the resolution cell corresponding to a scene pixel is assumed to be a DS according to section 2.1.1.

CCG signals are decomposed into two autarkic components $x_{k,i} = \mathfrak{Re}(x_{k,i}) + \mathrm{j}\,\mathfrak{Im}(x_{k,i})$. The real $\mathfrak{Re}(x_{k,i})$ and imaginary $\mathfrak{Im}(x_{k,i})$ parts are independent Gaussian distributed random variables with zero mean $\mathrm{E}\{\mathfrak{Re}(x_{k,i})\} = 0$ and $\mathrm{E}\{\mathfrak{Im}(x_{k,i})\} = 0$. Within each channel k, the standard deviation of the real part σ_{re} is the same as of the imaginary component σ_{im}, i.e., $\sigma_{\mathrm{re}} = \sigma_{\mathrm{im}} = \sigma_k$

$$\begin{aligned} \mathfrak{Re}(x_{k,i}) &\sim \mathcal{N}(0, \sigma_k) \\ \mathfrak{Im}(x_{k,i}) &\sim \mathcal{N}(0, \sigma_k). \end{aligned} \tag{2.1}$$

This also means that the amplitude is Rayleigh distributed $a_{k,i} \sim p_a(a_{k,i} \mid \sigma_k)$ with

$$p_a(a \mid \sigma) = \frac{a}{\sigma^2} \exp(-\frac{a^2}{2\sigma^2}). \tag{2.2}$$

The component σ_k is related to the channel signal power $I_k = 2\sigma_k^2$. However, the phase of a SAR sample is uniform distributed $\phi_{k,i} \sim \mathrm{U}(-\pi, \pi)$, i.e., the phase of a SAR scene alone is not usable. Practically, the random object phase $\phi_{k,i}^{(\mathrm{obj})}$ is the reason.

With SAR, the specific measured phase $\phi_{k,i}$ of scene k and pixel with index i is the superposition of the range distance $R_{k,i}$ related phase $\phi_{k,i}^{(\mathrm{R})} = \frac{4\pi}{\lambda} R_{k,i}$ which is the signal component, and the random object phase $\phi_{k,i}^{(\mathrm{obj})}$ as well as the unavoidable noise $\phi_{k,i}^{(\mathrm{noise})}$ introduced by the electronics, analog to digital conversion and approximations in the focusing

$$\phi_{k,i} = \phi_{k,i}^{(\mathrm{obj})} + \phi_{k,i}^{(\mathrm{R})} + \phi_{k,i}^{(\mathrm{noise})}. \tag{2.3}$$

InSAR combines two SAR scenes pixel by pixel, and ideally eliminates the random and not predictable object phase $\phi_{k,i}^{(\mathrm{obj})}$. For the following considerations, it is unimportant which two scenes are combined. Therefore, the scene index k is set to 1 and 2 for the

primary and secondary scenes, respectively. An interferogram sample z_i is the conjugate complex product of the primary and secondary SLC SAR scene pixel at index i

$$\begin{aligned} z_i &= |z_i|\, \mathrm{e}^{\mathrm{j}\varphi_i} \\ &= x_{1,i}\; x_{2,i}^{*}. \end{aligned} \tag{2.4}$$

The interferometric phase of the pixel is

$$\varphi_i = \arg\left(\mathrm{e}^{\mathrm{j}(\phi_{1,i}-\phi_{2,i})}\right). \tag{2.5}$$

For a single resolution cell on ground represented by a pixel in the interferogram, a 2×2 covariance matrix $\boldsymbol{\Sigma}$ describes the relation of the respective CCG processes $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. For this, $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ must be jointly circular Gaussian. In addition to the characteristics of (2.1), the conditions

$$\begin{aligned} \mathrm{E}\{\mathfrak{Re}(x_{1,i})\,\mathfrak{Re}(x_{2,i})\} &= \mathrm{E}\{\mathfrak{Im}(x_{1,i})\,\mathfrak{Im}(x_{2,i})\} \\ \mathrm{E}\{\mathfrak{Im}(x_{1,i})\,\mathfrak{Re}(x_{2,i})\} &= -\,\mathrm{E}\{\mathfrak{Re}(x_{1,i})\,\mathfrak{Im}(x_{2,i})\} \end{aligned} \tag{2.6}$$

must be met according to Lee et al. (1993) and Touzi et al. (1999). That this assumption is valid is accepted for distributed scatterer. The covariance matrix $\boldsymbol{\Sigma}$ contains the expected intensities on the diagonal and the covariances on the off-diagonal

$$\boldsymbol{\Sigma} = \begin{pmatrix} I_1 & \mathrm{Cov}(\boldsymbol{X}_1, \boldsymbol{X}_2) \\ \mathrm{Cov}(\boldsymbol{X}_1, \boldsymbol{X}_2)^{*} & I_2 \end{pmatrix}. \tag{2.7}$$

Basically, the covariance is defined for the random variables $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, i.e., for the primary and the secondary signal by

$$\mathrm{Cov}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \mathrm{E}\{\boldsymbol{X}_1 \boldsymbol{X}_2^{*}\} - \mathrm{E}\{\boldsymbol{X}_1\}\;\mathrm{E}\{\boldsymbol{X}_2^{*}\} \tag{2.8}$$

where $\mathrm{E}\{.\}$ is the expected value operator, i.e., the ensemble average. The linear correlation coefficient $\rho_{\boldsymbol{X}_1,\boldsymbol{X}_2}$ is defined by the normalization

$$\rho_{\boldsymbol{X}_1,\boldsymbol{X}_2} = \frac{\mathrm{Cov}(\boldsymbol{X}_1, \boldsymbol{X}_2)}{\sqrt{\mathrm{Var}(\boldsymbol{X}_1)}\,\sqrt{\mathrm{Var}(\boldsymbol{X}_2)}}. \tag{2.9}$$

For a CCG process, (2.8) simplifies with $\mathrm{E}\{\boldsymbol{X}_1\} = 0$ and $\mathrm{E}\{\boldsymbol{X}_2\} = 0$ and the complex coherence is

$$\gamma\, \mathrm{e}^{\mathrm{j}\varphi} = \frac{\mathrm{E}\{\boldsymbol{X}_1\; \boldsymbol{X}_2^{*}\}}{\sqrt{\mathrm{E}\{|\boldsymbol{X}_1|^2\}}\,\sqrt{\mathrm{E}\{|\boldsymbol{X}_2|^2\}}}. \tag{2.10}$$

The principle of ergodicity allows using the spatial mean instead of the ensemble average and gives the sample estimator of the coherence coefficient $\hat{\gamma}_{\mathrm{s}}$ based on N i.i.d. samples. In practice, a homogenous area is detected providing a spatially homogenous backscatter coefficient and a stationary phase signal (i.e., residual topography, deformation and atmospheric phase screen are compensated). Finally, the random processes are represented by the samples of each channel $\boldsymbol{X}_1 : \{x_{1,1}, x_{1,2}, \ldots, x_{1,N}\}$ and $\boldsymbol{X}_2 : \{x_{2,1}, x_{2,2}, \ldots, x_{2,N}\}$.

In operational systems, the sample estimator $\hat{\gamma}_\mathrm{s}$ is typically implemented to infer the coherence magnitude γ

$$\hat{\gamma}_\mathrm{s}\, \mathrm{e}^{\mathrm{j}\hat{\varphi}_\mathrm{s}} = \frac{\sum_{i=1}^{N} x_{1,i}\, x_{2,i}^{*}}{\sqrt{\sum_{i=1}^{N} |x_{1,i}|^2}\, \sqrt{\sum_{i=1}^{N} |x_{2,i}|^2}}. \tag{2.11}$$

Carter et al. (1973) as well as Touzi & Lopes (1996) show this is a random variable which is systematically biased toward higher coherence and provide the statistic of this sample estimate based on the conditional probability density $\mathrm{p}_{\hat{\gamma}_\mathrm{s}\,|\,\gamma}(\hat{\gamma}_\mathrm{s} \mid \gamma, N)$ visualized in Fig. 2.2. It depends on the number of samples N and the underlying coherence γ:

$$\mathrm{p}_{\hat{\gamma}_\mathrm{s}\,|\,\gamma}(\hat{\gamma}_\mathrm{s} \mid \gamma, N) = 2\hat{\gamma}_\mathrm{s} \left(1-\hat{\gamma}_\mathrm{s}^2\right)^{N-2} (N-1)(1-\gamma^2)^N\, {}_2\mathrm{F}_1\left(N, N; 1; \gamma^2\, \hat{\gamma}_\mathrm{s}^2\right). \tag{2.12}$$

In the equation above, ${}_2\mathrm{F}_1(.)$ is the Gaussian hypergeometric function. Carter et al. (1973) also provide the respective cumulative distribution function (CDF)

$$\mathrm{P}_{\hat{\gamma}_\mathrm{s}\,|\,\gamma}(\hat{\gamma}_\mathrm{s} \mid \gamma, N) = \hat{\gamma}_\mathrm{s}^2 \left(\frac{1-\gamma^2}{1-\gamma^2\hat{\gamma}_\mathrm{s}^2}\right)^{N} \sum_{k=0}^{N-2} \left(\frac{1-\hat{\gamma}_\mathrm{s}^2}{1-\gamma^2\hat{\gamma}_\mathrm{s}^2}\right)^{2} {}_2\mathrm{F}_1\left(-k, 1-N; 1; \gamma^2\, \hat{\gamma}_\mathrm{s}^2\right) \tag{2.13}$$

and the m-th raw moment

$$\mathrm{E}\{\hat{\gamma}_\mathrm{s}^m \mid \gamma, N\} = \left(1-\gamma^2\right)^N \frac{\Gamma(N)\Gamma(\frac{m}{2}+1)}{\Gamma(\frac{m}{2}+N)}\, {}_3\mathrm{F}_2\left(\frac{m}{2}+1, N, N; \frac{m}{2}+N, 1; \gamma^2\right). \tag{2.14}$$

In the equation above, ${}_3\mathrm{F}_2(.)$ is the generalized hypergeometric function. The expected value is computed from (2.14) with $m = 1$

$$\mathrm{E}\{\hat{\gamma}_\mathrm{s} \mid \gamma, N\} = \frac{\sqrt{\pi}\,(1-\gamma^2)^N\, \Gamma(N)\, {}_3\mathrm{F}_2\left(\frac{3}{2}, N, N; 1, N+\frac{1}{2}; \gamma^2\right)}{2\Gamma\left(N+\frac{1}{2}\right)} \tag{2.15}$$

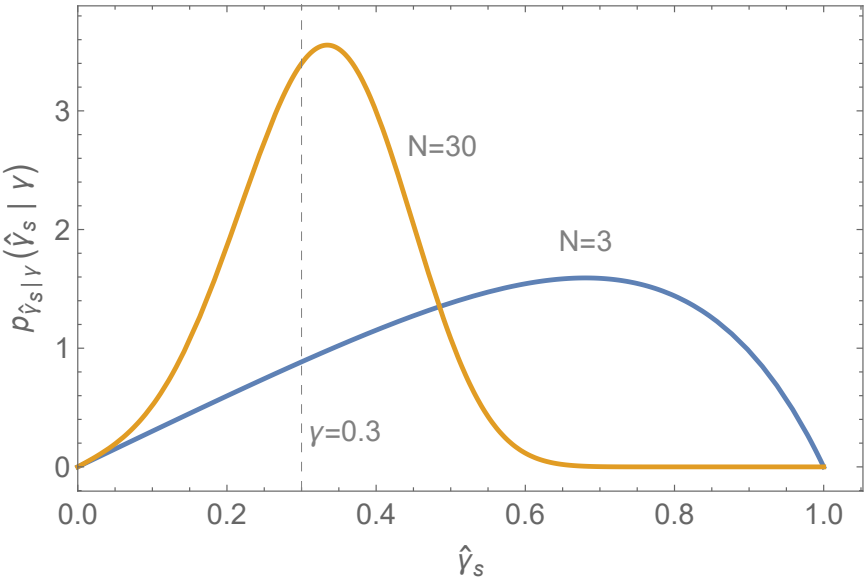


Figure 2.2: Example conditional probability density (2.12) with $\gamma = 0.3$ for $N = 3$ (blue) and $N = 30$ (orange).

and the second raw moment, i.e., $m = 2$ in (2.14), by

$$\mathrm{E}\{\hat{\gamma}_\mathrm{s}^2 \mid \gamma, N\} = \frac{(1-\gamma^2)^N \, \Gamma(N) \, {}_3\mathrm{F}_2(2, n, n; 1, n+1; \gamma^2)}{\Gamma(N+1)}. \tag{2.16}$$

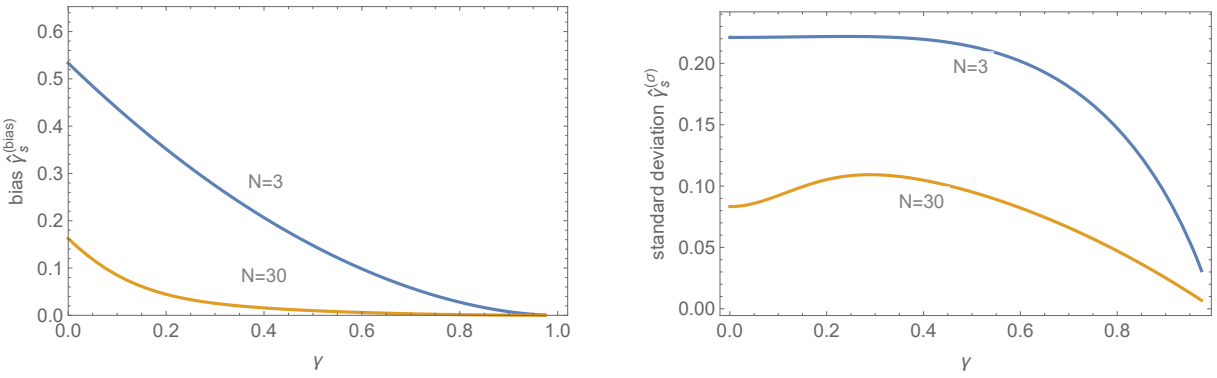
Carter et al. (1973) as well as Touzi et al. (1999) use the relation between the central moments and the raw moments to calculate the sample estimator bias $\gamma_\mathrm{s}^{(\mathrm{bias})}$

$$\gamma_\mathrm{s}^{(\mathrm{bias})} = \mathrm{E}\{\hat{\gamma}_\mathrm{s} \mid \gamma, N\} - \gamma \tag{2.17}$$

and the sample estimation standard deviation $\gamma_\mathrm{s}^{(\sigma)}$

$$\begin{aligned} \gamma_\mathrm{s}^{(\sigma)} &= \sqrt{\mathrm{E}\{\hat{\gamma}_\mathrm{s}^2 \mid \gamma, N\} - \mathrm{E}\{\hat{\gamma}_\mathrm{s} \mid \gamma, N\}^2} \\ &= \left(\frac{(1-\gamma^2)^N \, \Gamma(N) \, {}_3\mathrm{F}_2(2, N, N; 1, N+1; \gamma^2)}{\Gamma(N+1)} \right. \\ &\qquad \left. - \frac{\pi \, (1-\gamma^2)^{2N} \, \Gamma(N)^2 \, {}_3\mathrm{F}_2\left(\frac{3}{2}, N, N; 1, N+\frac{1}{2}; \gamma^2\right)^2}{4 \, \Gamma\left(N+\frac{1}{2}\right)^2} \right)^{\frac{1}{2}}. \end{aligned} \tag{2.18}$$

Figure 2.3 visualizes examples of the sample estimator bias (2.17) and standard deviation (2.18). By Fig. 2.3a, it is clear, the sample estimator is asymptotically unbiased, i.e., for large N and/or high coherence γ. By additionally looking at Fig. 2.3b, it can be seen that the sample estimator is consistent, meaning that it estimates more accurately as the sample size increases. Efficiency is a criterion for the best possible estimator. An estimator is efficient if it achieves the Cramér-Rao Lower Bound (CRLB) or, formulated for estimators that are biased, achieves the best possible mean squared error (MSE) or RMSE. This is certainly the case if both the bias $\gamma_s^{(\mathrm{bias})}$ (2.17) and the variance $(\gamma_s^{(\sigma)})^2$ (2.18) are zero. It can be shown that $\lim_{\gamma \to 1} \gamma_s^{(\mathrm{bias})} = 0$ and $\lim_{\gamma \to 1} \gamma_s^{(\sigma)} = 0$. Therefore, the sample estimator is efficient for an underlying coherence $\gamma = 1$ and all sample sizes $N \geq 2$. The limits as N goes to infinity can only be calculated for trivial γ (e.g., $\gamma = 0$ or $\gamma = 1$). But it is a feature of maximum likelihood estimates (MLEs) that they are,



(a) Bias. *(b)* Standard deviation.

Figure 2.3: Characteristics of the sample estimation $\hat{\gamma}_\mathrm{s}$ for $N = 3$ (blue) and $N = 30$ (orange).

with respect to sample size, asymptotically unbiased and have an asymptotically minimal variance. Therefore, the sample estimator is efficient when N is infinity and/or $\gamma = 1$. However, both cases of $\gamma = 1$ and $N = \infty$ are rather theoretical. In practice, these results mean that other estimators cannot be better for these extreme configurations, and it becomes more difficult for other estimators the closer one gets to these extremes. I.e., N is large and/or the coherence is large.

Olkin & Pratt (1958) formulate the condition to construct an unbiased estimator $G(\hat{\gamma}_\mathrm{s})$ based on the sample estimate (2.11)

$$\mathrm{E}\{G(\hat{\gamma}_\mathrm{s})\} = \gamma. \tag{2.19}$$

Based on the basic approach

$$\mathrm{E}\{G(\hat{\gamma}_\mathrm{s})\} = \int_0^1 G(\hat{\gamma}_\mathrm{s})\, \mathrm{p}_{\hat{\gamma}_\mathrm{s}}(\hat{\gamma}_\mathrm{s} \mid \gamma, N)\, d\hat{\gamma}_\mathrm{s}, \tag{2.20}$$

Touzi et al. (1999) prove that $G(\hat{\gamma}_\mathrm{s})$ cannot be constructed. As a consequence, all coherence estimators are biased, and importantly, the conventional CRLB, i.e. the inverse Fisher information matrix (FIM), is not relevant for the variance of such estimators.

2.1.3 InSAR Data Statistic and Coherence Applications

The probability density functions (PDFs) below are independent from the pixel position i, i.e., they are valid for each jointly CCG sample vector and the indices are associated with the SLC acquisition k. For this reason, the pixel index i is not present in the variable names. Generally, for N_SLC scenes, the conditional PDF with an underlying (given) complex covariance matrix $\boldsymbol{\Sigma}$ of a zero mean complex N_SLC-variate Gaussian random variable $\boldsymbol{x} = \boldsymbol{X}_{*,i} = \begin{pmatrix} x_{1,i} & \dots & x_{k,i} & \dots & x_{N_\mathrm{SLC},i} \end{pmatrix}^T$ is

$$\mathrm{p}_\mathrm{CCG}(\boldsymbol{x} \mid \boldsymbol{\Sigma}) = \frac{1}{\pi^{N_\mathrm{SLC}} \det(\boldsymbol{\Sigma})} \exp\left(-\boldsymbol{x}^H \boldsymbol{\Sigma}^{-1} \boldsymbol{x}\right). \tag{2.21}$$

Based on the equation above, the joint PDF of an InSAR data pair, i.e., $N_\mathrm{SLC} = 2$, is published by Touzi & Lopes (1996) as a function of the primary scene amplitude a_1 and the secondary scene amplitude a_2 as well as the interferometric phase difference $\varphi_{1,2} = \phi_1 - \phi_2$ given the underlying coherence γ and true interferometric phase φ

$$\begin{aligned} &\mathrm{p}_{\mathrm{InSAR}_1}(a_1, a_2, \varphi_{1,2} \mid \gamma, \varphi, I_1, I_2) = \\ &\qquad \frac{2\, a_1\, a_2}{\pi\, I_1\, I_2\, (1-\gamma^2)} \exp\left(-\frac{I_2\, a_1^2 + I_1\, a_2^2 - 2 a_1\, a_2 \sqrt{I_1\, I_2}\, \gamma\, \cos(\varphi_{1,2} - \varphi)}{I_1\, I_2\, (1-\gamma^2)}\right). \end{aligned} \tag{2.22}$$

The equation above is the general joint PDF of two correlated complex samples with individual expected intensities I_1 and I_2 and is valid for $a_1 \geq 0, a_2 \geq 0, I_1 > 0, I_2 > 0$ and $0 \leq \gamma < 1$. The phase difference is relevant only, because the object phase is completely

random and does not contribute to the statistic. A more specific joint PDF with given parameter $\sigma = \sqrt{I_1/2} = \sqrt{I_2/2}$ is published by Goodman (1975)

$$\mathrm{p}_{\mathrm{InSAR}_2}(a_1, a_2, \varphi_{1,2} \mid \gamma, \varphi, \sigma) = \frac{a_1\, a_2}{2\,\pi\,\sigma^4\,(1-\gamma^2)} \exp\left(-\frac{a_1^2 + a_2^2 - 2a_1\, a_2\, \gamma\, \cos(\varphi_{1,2} - \varphi)}{2\,\sigma^2\,(1-\gamma^2)}\right). \tag{2.23}$$

This equation assumes equal channel intensities $I_1 = I_2$ and uses the standard deviation σ of the real and imaginary components as parameter. Straight forward to implement estimators for the MLE are known for (2.22) and (2.23). According to Touzi et al. (1999), the sample coherence magnitude $\hat{\gamma}_\mathrm{s}$ in (2.22) is the MLE of the coherence magnitude γ. Seymour & Cumming (1994) developed the MLE based on (2.23).

Integrating (2.22) over a_1 and a_2 provides the marginal PDF of the measured interferometric phase difference $\varphi_{1,2}$. It is published by Goodman (1975), Lee et al. (1993), Bamler & Hartl (1998) and in more detail by Goodman (2007)

$$\mathrm{p}_{\varphi_{1,2}}(\varphi_{1,2} \mid \gamma, \varphi) = \frac{(1-\gamma^2)\left(\sqrt{1-\beta^2} + \pi\beta - \beta\arccos(\beta)\right)}{2\pi\,(1-\beta^2)^{3/2}}. \tag{2.24}$$

In the equation above, the term $\beta = \gamma\cos(\varphi_{1,2} - \varphi)$ needs to be substituted. Finally, the expected value

$$\mathrm{E}\{\varphi_{1,2} \mid \gamma, \varphi\} = \varphi \tag{2.25}$$

and standard deviation of the interferometric phase

$$\sigma_{\varphi_{1,2}} = \left(\frac{\pi^2}{3} - \pi\arcsin(\gamma) + \arcsin^2(\gamma) - \frac{1}{2}\,\mathrm{Li}_2(\gamma^2)\right)^{1/2} \tag{2.26}$$

result, where $\mathrm{Li}_2(.)$ is the dilogarithm function. The last equation clearly shows that the coherence γ alone determines the standard deviation of the InSAR phase $\varphi_{1,2}$ and thus its precision. However, the multilook statistic, i.e., for the InSAR phase sample estimate

$$\hat{\varphi}_\mathrm{s} = \arg\left(\sum_{i=1}^{N} x_{1,i}\, x_{2,i}^*\right), \tag{2.27}$$

is more relevant in practice. Lee et al. (1993) have published the respective PDF

$$\mathrm{p}_{\hat{\varphi}_\mathrm{s}}(\hat{\varphi}_\mathrm{s} \mid \gamma, \varphi, N) = \frac{(1-\gamma^2)^N}{2\pi}\left({}_2\mathrm{F}_1\left(1, N; \frac{1}{2}; \beta^2\right) + \frac{\sqrt{\pi}\,\Gamma\left(N+\frac{1}{2}\right)\beta}{\Gamma(N)\,(1-\beta^2)^{N+\frac{1}{2}}}\right). \tag{2.28}$$

In the equation above, the term $\beta = \gamma\cos(\hat{\varphi}_\mathrm{s} - \varphi)$ needs to be substituted. Unfortunately, there is no closed-form solution for the standard deviation of the multilook phase $\sigma_{\hat{\varphi}_\mathrm{s}} = \sqrt{\mathrm{E}\{(\hat{\varphi}_\mathrm{s} - \varphi)^2\}}$. To describe the exact functional relationship between $\sigma_{\hat{\varphi}_\mathrm{s}}$ and γ, numerical integration is used (Lee et al. 1993, Touzi & Lopes 1996, Bamler & Hartl 1998). Rosen et al. (2000) suggest using the Cramer–Rao bound as a reasonable approximation

$$\sigma_{\hat{\varphi}_\mathrm{s}} \approx \frac{1}{\sqrt{2N}}\frac{\sqrt{1-\gamma^2}}{\gamma} \tag{2.29}$$

when $N > 4$.

The applications of coherence are based on its relationship to the InSAR phase standard deviation $\sigma_{\hat{\varphi}_\mathrm{s}}$ described above. For instance, for DEM mapping, the error propagation from phase standard deviation to topography height standard deviation $\sigma_{\hat{h}}$ is

$$\sigma_{\hat{h}} = \sqrt{k_z^2\,\sigma_{\hat{\varphi}_\mathrm{s}}^2}. \tag{2.30}$$

k_z is the vertical wavenumber or, more precisely, the phase to height conversion factor which is for bistatic single-pass interferometry

$$k_z = \frac{2\pi}{\lambda}\frac{B_\mathrm{eff}}{R_0\sin(\theta_0)}. \tag{2.31}$$

It depends on the observation geometry, i.e., the effective baseline B_eff, the range distance from sensor to the ground R_0, the radar wavelength λ and the local incidence angle θ_0. First, the DEM precision $\sigma_{\hat{h}}$ can be provided as an operational quality data layer based on the estimated coherence, i.e., based on the actual data quality. Second, because $\sigma_{\hat{\varphi}_\mathrm{s}}$ can be tuned by the number of averaged samples N, the spatial resolution and the precision of the height monitoring can be adjusted according to the requirements. Furthermore, in operational processing, the coherence is an essential layer to support phase unwrapping (Derauw 1995).

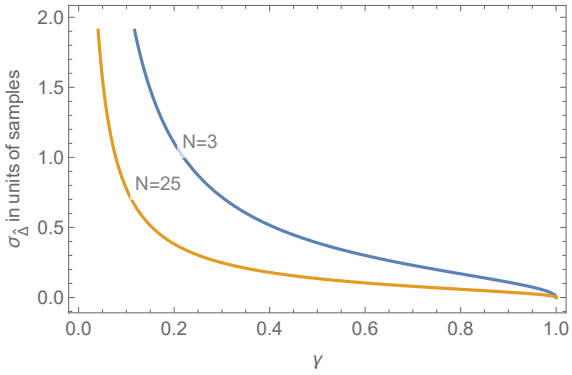
Another application of the coherence magnitude is the estimation of the coregistration accuracy depending on the underlying coherence. In InSAR systems, the complex cross correlation is used to estimate the mutual shift $\hat{\Delta}$ for the coregistration of the primary and secondary scene. According to Bamler (2000), the standard deviation of this shift estimate $\sigma_{\hat{\Delta}}$ for homogeneous image patches in units of samples is

$$\sigma_{\hat{\Delta}} = \sqrt{\frac{3}{2N}}\frac{\sqrt{1-\gamma^2}}{\pi\gamma}r^{3/2}. \tag{2.32}$$

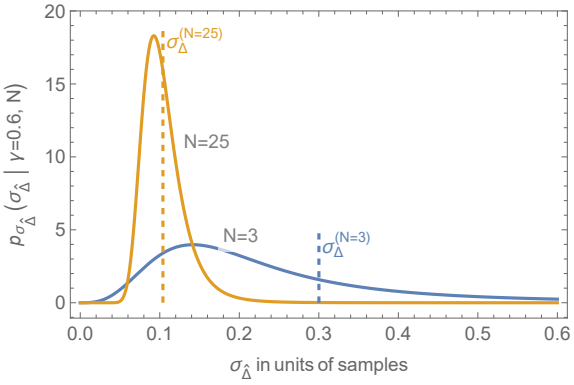
r is the oversampling of the data. Figure 2.4a visualizes this functional relationship. In order to provide the actual coregistration accuracy during InSAR processing, the coherence needs to be estimated from the data. Accordingly, the coherence γ is to be substituted by the sample estimate $\hat{\gamma}_\mathrm{s}$ with the PDF (2.12), i.e., $\hat{\gamma}_\mathrm{s} \sim \mathrm{p}_{\hat{\gamma}_\mathrm{s}\,|\,\gamma}(\hat{\gamma}_\mathrm{s} \mid \gamma, N)$. Applying the transformation of random variables, the PDF of the standard deviation of the shift estimate

$$\begin{aligned}\mathrm{p}_{\sigma_{\hat{\Delta}}}(\sigma_{\hat{\Delta}} \mid \gamma, N) =& \frac{3(N-1)r^3}{\pi^2 N\sigma_{\hat{\Delta}}^3}\left(\left(\gamma^2-1\right)\left(\frac{3r^3}{2\pi^2 N\sigma_{\hat{\Delta}}^2+3r^3}-1\right)\right)^N \\ &\times {}_2\mathrm{F}_1\left(N, N; 1; \frac{3r^3\gamma^2}{3r^3+2N\pi^2\sigma_{\hat{\Delta}}^2}\right)\end{aligned} \tag{2.33}$$

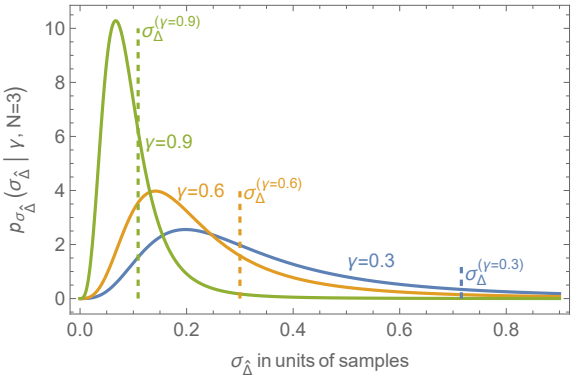
is obtained. This PDF is visualized in Fig. 2.4b for $N = 3$ and $N = 25$. In addition, the true coregistration standard deviations are shown as dashed lines. As $\hat{\gamma}_\mathrm{s}$ has a bias and a variance, the accuracy estimate $\sigma_{\hat{\Delta}}$ is also subject to bias and dispersion based on the



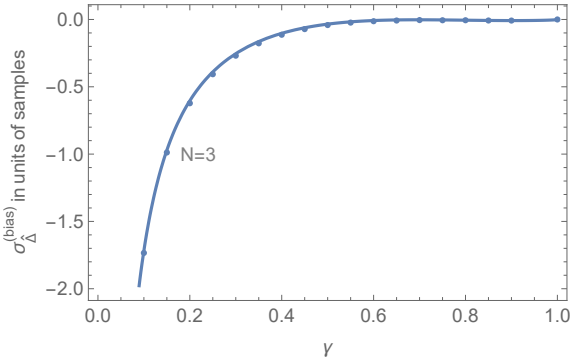
(a) $\sigma_{\hat{\Delta}}$ depending on the underlying coherence.



(b) $\sigma_{\hat{\Delta}}$-PDF resulting from sample estimator.



(c) Example $\sigma_{\hat{\Delta}}$-PDFs for $N = 3$.



(d) Bias of $\sigma_{\hat{\Delta}}$ resulting from sample estimator.

Figure 2.4: The estimated coregistration standard deviation $\sigma_{\hat{\Delta}}$ depends on the characteristics of the coherence estimator used. (b), (c) and (d) are examples for the sample estimator.

fundamental coherence. Figure 2.4c visualizes as an example for $N = 3$ the estimation likelihoods depending on the true coherence. For specific N, the bias can be provided as a closed-form expression, e.g., for $N = 3$ it is

$$\sigma_{\hat{\Delta}}^{(\text{bias})} = \frac{\sqrt{2-2\gamma^2}\, r^{3/2}}{64\pi\gamma}\left(3\pi\gamma\left(3\gamma^4 - 8\gamma^2 + 8\right) - 32\right). \tag{2.34}$$

The corresponding graph is visualized in Fig. 2.4d together with the verification simulation (dots). Due to the characteristics of the sample estimator, however, the estimate (2.32) is not always completely precise, especially with low coherence and few data points.

There is now a very important application for coherence magnitude in InSAR time series analysis for ground motion monitoring. The respective methods have been published by Guarnieri & Tebaldini (2008), by Ferretti et al. (2011) with SqueeSAR, by Perissin & Wang (2012) and by Fornaro et al. (2015) with CAESAR. These methods estimate long time span coherent InSAR phases $\arg(\hat{\boldsymbol{x}}_{\text{PS}})$ from DSs which originally decorrelate over time. For this purpose, a vector $\boldsymbol{x}_{\text{PS}} = \left(\mathrm{e}^{\mathrm{j}\phi_1} \;\ldots\; \mathrm{e}^{\mathrm{j}\phi_k} \;\ldots\; \mathrm{e}^{\mathrm{j}\phi_{N_{\text{SLC}}}}\right)^T$ is analyzed to optimally represent the data matrix $\boldsymbol{X}$ of the statistically homogenous pixels (SHP) with respect to the phases. Such estimates can be used directly in a PSI processing complementing PSs and therefore make a significant contribution to increasing the measurement

density outside urban areas. The basis is the detection of SHP, which provides i.i.d. samples per channel and allows a statistical analysis by exploiting the principle of ergodicity. Ferretti et al. (2011) as well as Parizzi & Brcic (2011) propose amplitude-based statistical goodness-of-fit tests to group i.i.d. pixels, i.e., the two-sample Kolmogorov-Smirnov, the Kullback-Leibler divergence, the Anderson-Darling, and the Generalized Likelihood Ratio tests. As (2.10) shows, it is an inherent principle that coherence is related to only two channels. The reason is, that the bivariate complex Gaussian distribution results in the single complex coherence parameter $\gamma \exp(\mathrm{j}\varphi)$ as the statistic parameter. For InSAR long time series analysis based on zero mean complex N_{SLC}-variate Gaussian random variables, the signal description can be extended to a coherency matrix $\boldsymbol{T}$. Each entry T_{k_1,k_2} at index k_1, k_2 in $\boldsymbol{T}$ corresponds to the respective complex coherence $T_{k_1,k_2} = \gamma_{k_1,k_2} \exp(\mathrm{j}\varphi_{k_1,k_2})$. Now, the task remains to estimate each entry $\hat{T}_{k_1,k_2} = \hat{\gamma}_{k_1,k_2} \exp(\mathrm{j}\hat{\varphi}_{k_1,k_2})$ in $\boldsymbol{T}$ for $k_1 = 2 \ldots N_{\mathrm{SLC}}$ and $k_2 = k_1 + 1 \ldots N_{\mathrm{SLC}}$. Two methods are presented to demonstrate that the coherence matrix $\boldsymbol{T}$ and as a consequence coherence estimation is fundamental for DS based InSAR time series estimation.

Ferretti et al. (2011) published the MLE based on the likelihood function formed from products of (2.21). Finally, the nonlinear estimation

$$\begin{aligned} \hat{\boldsymbol{x}}_{\mathrm{PS}} &= \underset{\boldsymbol{x}_{\mathrm{PS}}}{\operatorname{argmax}} \left(\boldsymbol{x}_{\mathrm{PS}}^H \left(\left| \hat{\boldsymbol{T}} \right|^{-1} \circ \hat{\boldsymbol{T}} \right) \boldsymbol{x}_{\mathrm{PS}} \right) \\ &\text{subject to } \ \phi_1 = 0. \end{aligned} \tag{2.35}$$

over the parameter vector $\boldsymbol{x}_{\mathrm{PS}}$ provides the solution $\arg(\hat{\boldsymbol{x}}_{\mathrm{PS}})$. The matrix operation $\circ$ represents the element-wise (i.e., the Hadamard) product.

Fornaro et al. (2015) published the phase estimation based on the Eigenvalue Decomposition (EVD) of the covariance matrix $\boldsymbol{\Sigma}$. PS like InSAR phases are estimated by extracting the phases of the eigenvector associated with the largest eigenvalue from the estimated covariance matrix

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N} \boldsymbol{X} \boldsymbol{X}^H. \tag{2.36}$$

According to Cao et al. (2015), the EVD can also be implemented on the coherence matrix $\boldsymbol{T}$ instead of the covariance matrix. This implies that the SHP data only shows one scattering mechanism. The conversion from a covariance matrix into a coherence matrix is

$$\hat{\boldsymbol{T}} = \hat{\boldsymbol{\Sigma}} \oslash \left(\hat{\boldsymbol{a}} \hat{\boldsymbol{a}}^T \right). \tag{2.37}$$

The symbol $\oslash$ is the Hadamard division, i.e., the matrix element-wise division. The vector $\hat{\boldsymbol{a}}$ consists of the square roots of the expected intensities I_k of each channel k and can be extracted from the covariance matrix using the identity matrix $\mathbb{I}$ and a vector of ones $\mathbf{1}$

$$\hat{\boldsymbol{a}} = \sqrt{\left(\hat{\boldsymbol{\Sigma}} \circ \mathbb{I} \right) \mathbf{1}}. \tag{2.38}$$

$\boldsymbol{T}$ is a Hermitian matrix and its eigendecomposition is the sum of outer products of its eigenvectors $\boldsymbol{u}_i$, weighted by their eigenvalues λ_i

$$\boldsymbol{T} = \sum_{i=1}^{N_{\mathrm{SLC}}} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^H. \tag{2.39}$$

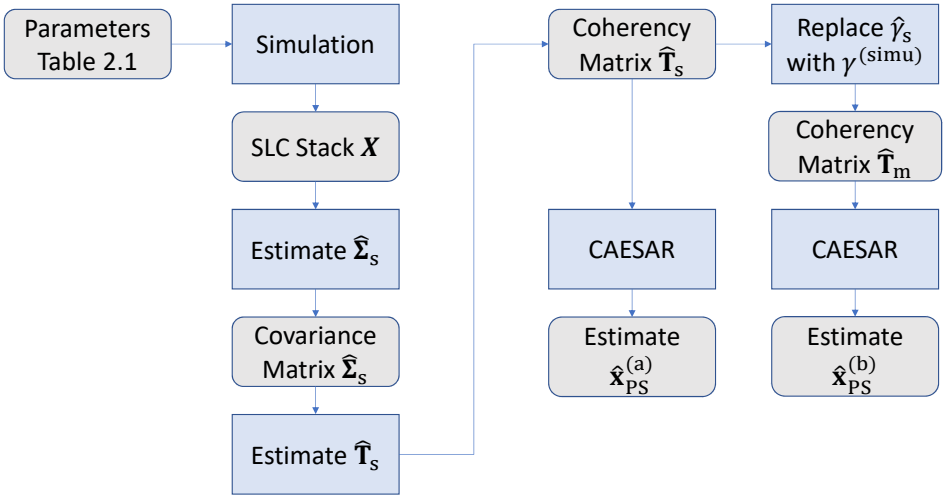


Figure 2.5: Principle of experiment.

Practically, the estimated coherence matrix $\hat{\boldsymbol{T}}$ is used and the method solves for

$$\begin{aligned} \hat{\boldsymbol{x}}_{\mathrm{PS}} &= \underset{\boldsymbol{y}_{\mathrm{PS}}}{\operatorname{argmax}} \left(\boldsymbol{y}_{\mathrm{PS}}^H \hat{\boldsymbol{T}} \boldsymbol{y}_{\mathrm{PS}} \right) \\ \text{subject to } & \boldsymbol{y}_{\mathrm{PS}}^H \boldsymbol{y}_{\mathrm{PS}} = 1 \end{aligned} \tag{2.40}$$

with $\boldsymbol{y}_{\mathrm{PS}} = \left(a_1 \mathrm{e}^{\mathrm{j}\phi_1} \quad \dots \quad a_k \mathrm{e}^{\mathrm{j}\phi_k} \quad \dots \quad a_{N_{\mathrm{SLC}}} \mathrm{e}^{\mathrm{j}\phi_{N_{\mathrm{SLC}}}} \right)^T$. However, the solution is computed directly by $\boldsymbol{u}_1$ without the need for a nonlinear solver for (2.40). Ferretti et al. (2011) mention in equation 8 this approach in the framework of SqueeSAR.

In the applications above, knowledge of the true coherence γ is originally assumed. In practice, however, the estimated $\hat{\gamma}$ needs to be used instead. It is clear that in all mentioned applications error propagation occurs due to uncertainties in $\hat{\gamma}$. The influence of the estimation uncertainty of the coherences $\hat{\gamma}_{k_1,k_2}$ on the phases estimated by CAESAR is demonstrated by an experiment. The principle is visualized in Fig. 2.5. The experiment is based on a data simulation with the parameters of Table 2.1. Rocca (2007) and Zheng

parameter and value	unit	description
$N_{\mathrm{SLC}} = 10$		number of SLC scenes, i.e., channels
$N = 10$		number of i.i.d. samples in each scene
$\tau = 40$	day	characteristic decorrelation time
$\gamma_\infty = 0$		long-term coherence
$T = 12$	day	repeat cycle of sensor
$\phi_k \sim U(-\pi, \pi)$	rad	simulated phase of scene k
$N_{\mathrm{simu}} = 1\,000\,000$		number of simulations

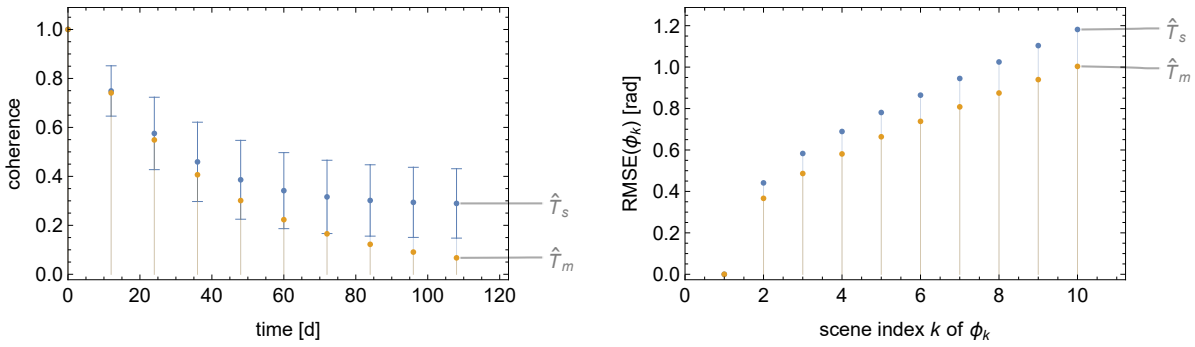
Table 2.1: Parameters to show how coherence magnitude impacts CAESAR phase estimation.

et al. (2020) describe a temporal decorrelation law of DSs

$$\gamma_{\text{model}}(t) = \gamma_\infty + (1 - \gamma_\infty) \exp(-\frac{t}{\tau}). \tag{2.41}$$

γ_∞ is the long-term coherence and τ represents the characteristic decorrelation time, which means that after a time t of 3τ, the modeled coherence has approximately decayed to the long-term coherence, i.e., $\gamma_{\text{model}}(3\tau) \approx \gamma_\infty$. With this model, SLC scenes are simulated and the data are represented by the matrix $\boldsymbol{X}$. Next, the sample covariance matrix $\hat{\boldsymbol{\Sigma}}_{\text{s}}$ is estimated according to (2.36), and (2.37) is used to calculate the coherence matrix $\hat{\boldsymbol{T}}_{\text{s}}$. This matrix is composed of complex sample coherence estimates $\hat{\gamma}_{k_1,k_2} \exp(\text{j}\hat{\varphi}_{k_1,k_2})$. On the one hand, this matrix is used directly for CAESAR estimation, resulting in the PS like InSAR phases $\arg(\hat{\boldsymbol{x}}_{\text{PS}}^{(\text{a})})$. On the other hand, it is modified. The estimated coherence magnitudes $\hat{\gamma}_{k_1,k_2}$ are replaced by the true values from the simulation $\gamma_{k_1,k_2}^{(\text{simu})}$ and the estimated phases $\hat{\varphi}_{k_1,k_2}$ are retained. The modified coherency matrix $\hat{\boldsymbol{T}}_{\text{m}}$ is also used for the CAESAR estimation resulting in $\arg(\hat{\boldsymbol{x}}_{\text{PS}}^{(\text{b})})$.

The experiment is evaluated using Fig. 2.6. Figure 2.6a shows the bias and standard deviation of the coherence estimates in $\hat{\boldsymbol{T}}_{\text{s}}$ (blue plot) and the true, i.e., simulated by (2.41), coherence (orange plot). In Figure 2.6b can be seen that the phase of the first scene is set to zero as a reference and therefore has no RMSE. Also, the more accurate the coherence is, the more accurate is the estimation accuracy of the phases. The RMSE of the phases $\arg(\hat{\boldsymbol{x}}_{\text{PS}}^{(\text{a})})$ (blue plot) is always larger than that of the phases $\arg(\hat{\boldsymbol{x}}_{\text{PS}}^{(\text{b})})$ and grows faster with increasing time interval (i.e., index number k). It is noteworthy that the improved PS like phase estimate results only from the coherence magnitude, since $\hat{\boldsymbol{T}}_{\text{s}}$ and $\hat{\boldsymbol{T}}_{\text{m}}$ contain the same phase estimates. Finally, the experiment confirms that the estimation accuracy of the DS estimation methods can be significantly improved by the accuracy of the coherence estimation.



(a) Simulated (orange) and estimated (blue) coherence. The error bar is one standard deviation.

(b) RMSE of phases from simulated (orange) and estimated (blue) coherence magnitude.

Figure 2.6: Estimation precision of coherences and phases from CAESAR. Orange: true coherence used. Blue: sample coherence estimates used.

2.2 State of the Art in Coherence Magnitude Estimation

Coherence magnitude estimation has been an active research field for decades. However, the main focus was on bias mitigation. In operational systems, still the sample coherence magnitude estimator is typically implemented.

Touzi et al. (1999) investigated two methods for unbiased estimation of the coherence magnitude. Both assume a stationary coherence in an area larger than the coherence estimation window. First, the coherence magnitude from a coherence magnitude map (CMM) is considered. It allows spatially averaging the absolute value of L independent sample coherence estimates from N samples $\hat{\gamma}_{s,i}$ from (2.11), i.e.,

$$\hat{\gamma}_{\mathrm{CMM}}^{(L)} = \frac{1}{L}\sum_{i=1}^{L}\hat{\gamma}_{\mathrm{s},i}. \tag{2.42}$$

Touzi et al. (1999) provide an example with $N = 4$ and $L = 1000$. $\hat{\gamma}_{\mathrm{CMM}}^{(L)}$ tends to be distributed normally about $\mathrm{E}\{\hat{\gamma}_{\mathrm{s}} \mid \gamma, N = 4\}$ given by (2.15) with standard deviation $\gamma_{\mathrm{s}}^{(\sigma)}/\sqrt{L}$ based on (2.18). The low standard deviation allows to invert (2.15) using a table. Touzi et al. (1999) state that for $\gamma < 0.1$ this bias corrected estimate is better than the sample estimate (2.11) based on all samples in the overall area. Second, the coherence estimation from a complex coherence map (CCM) is proposed with

$$\hat{\gamma}_{\mathrm{CCM}}^{(L)} = \frac{1}{L}\left|\sum_{i=1}^{L}\hat{\gamma}_{\mathrm{s},i}\exp(\mathrm{j}\hat{\varphi}_{\mathrm{s},i})\right|. \tag{2.43}$$

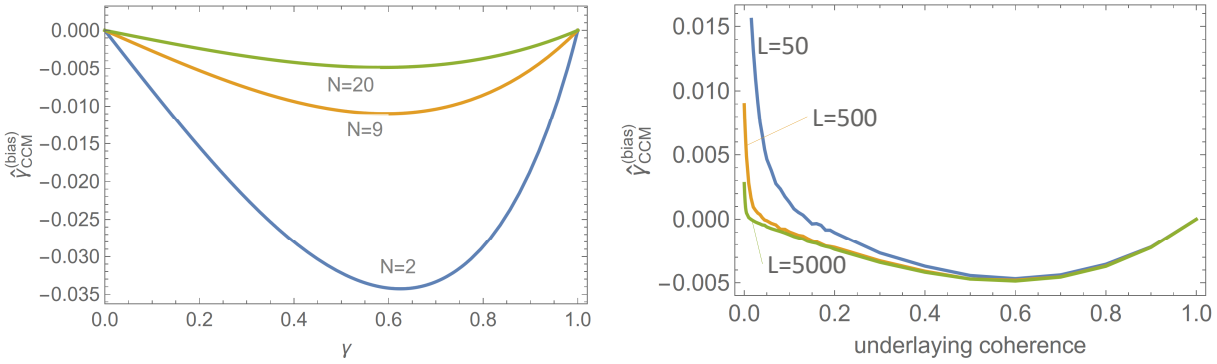
It is shown that a nearly unbiased coherence estimate results from spatially averaging the complex coherence map of sufficiently coarse resolution, e.g., $N = 20$. The proof starts with the joint PDF of the complex coherence (Touzi et al. 1999, equation 48)

$$\begin{aligned}\mathrm{p}_{\hat{\gamma}_{\mathrm{s}},\hat{\varphi}_{\mathrm{s}}}(\hat{\gamma}_{\mathrm{s}},\hat{\varphi}_{\mathrm{s}} \mid \gamma,\varphi,N) &= \frac{(1-\gamma^2)^N\hat{\gamma}_{\mathrm{s}}(1-\hat{\gamma}_{\mathrm{s}}^2)^{N-2}}{\pi\Gamma(N)\Gamma(N-1)}\\ &\cdot\left[\Gamma^2(N)\,{}_2\mathrm{F}_1\Big(N,N;\frac{1}{2};\hat{\gamma}_{\mathrm{s}}^2\gamma^2\cos^2(\hat{\varphi}_{\mathrm{s}}-\varphi)\Big)\right.\\ &\left.+2\Gamma^2(N+\frac{1}{2})\hat{\gamma}_{\mathrm{s}}\gamma\cos(\hat{\varphi}_{\mathrm{s}}-\varphi)\,{}_2\mathrm{F}_1\Big(N+\frac{1}{2},N+\frac{1}{2};\frac{3}{2};\hat{\gamma}_{\mathrm{s}}^2\gamma^2\cos^2(\hat{\varphi}_{\mathrm{s}}-\varphi)\Big)\right].\end{aligned} \tag{2.44}$$

The expected value for the complex coherence $\hat{\gamma}_{\mathrm{s}}\exp(\mathrm{j}\hat{\varphi}_{\mathrm{s}})$ is

$$\begin{aligned}&\mathrm{E}\{\hat{\gamma}_{\mathrm{s}}\exp(\mathrm{j}\hat{\varphi}_{\mathrm{s}}) \mid \gamma,\varphi,N\}\\ &= \int_0^1\int_0^{2\pi}\hat{\gamma}_{\mathrm{s}}\exp(\mathrm{j}\hat{\varphi}_{\mathrm{s}})\,\mathrm{p}_{\hat{\gamma}_{\mathrm{s}},\hat{\varphi}_{\mathrm{s}}}(\hat{\gamma}_{\mathrm{s}},\hat{\varphi}_{\mathrm{s}} \mid \gamma,\varphi_0,N)\,\mathrm{d}\hat{\varphi}_{\mathrm{s}}\mathrm{d}\hat{\gamma}_{\mathrm{s}}\\ &= \frac{\Gamma^2(N+\frac{1}{2})}{\Gamma(N)\Gamma(N+1)}{}_2\mathrm{F}_1\Big(N+\frac{1}{2},N+\frac{1}{2};N+1;\gamma^2\Big)\,\gamma(1-\gamma^2)^N\exp(\mathrm{j}\varphi).\end{aligned} \tag{2.45}$$

Since $\mathrm{E}\{\hat{\gamma}_{\mathrm{s}}\exp(\mathrm{j}\hat{\varphi}_{\mathrm{s}}) \mid \gamma,\varphi,N\} \neq 0$, the estimator is biased. However, the bias has an interesting characteristic which is visualized in Fig. 2.7. Figure 2.7a demonstrates that



(a) Theoretical bias (2.45) depending on the sample size N.

(b) Bias depending on the sample size L from simulations with $N = 20$.

Figure 2.7: Bias characteristic of CCM estimator.

the bias reduces increasing the partial sample size N and is negative, i.e., the coherence magnitude is underestimated. What is remarkable about this figure is that, in contrast to the conventional sample estimator, small coherences are without bias. In order to realistically estimate the performance, one must consider the dependence on the number L of averaged complex coherences. Figure 2.7b shows that the theoretical bias can only be achieved with very large L and in practice, small coherences are estimated worse compared to Fig. 2.7a. For an estimate, a total of $T = NL$ i.i.d. samples are required and for the example in Fig. 2.7b with $N = 20$, $T = 1000$ for $L = 50$ (blue graph), $T = 10\,000$ for $L = 500$ (orange graph) and $T = 100\,000$ i.i.d. samples for $L = 5000$ (green graph).

Another work on bias removal for coherence magnitude estimation has been published by Abdelfattah & Nicolas (2006). Similar to the CMM and CCM of Touzi et al. (1999), it also assumes a stationary coherence in an area larger than the coherence estimation window. In this work, the coherence is computed from the logarithm of the sample coherence, i.e., a logarithmic coherence magnitude map (LCMM)

$$\hat{\gamma}_{\mathrm{LCMM}}^{(L)} = \frac{1}{L}\sum_{i=1}^{L} \log\left(\hat{\gamma}_{\mathrm{s},i}\right). \tag{2.46}$$

The principle of second kind statistic allows to evaluate the bias of this estimator and to compensate the biased estimate. Practically, the expected value of the estimate (2.46) is computed using (2.12) from the first log-moment

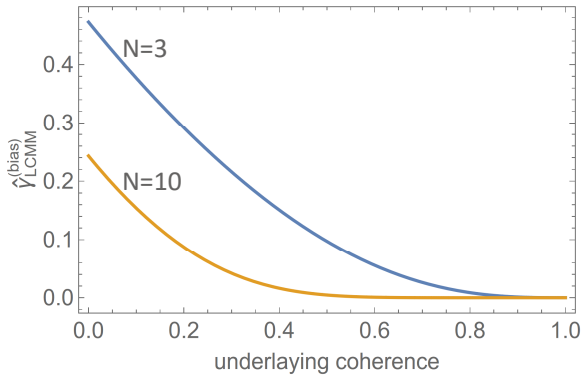
$$\underline{m}_1 = \int_0^1 \log\left(\hat{\gamma}_{\mathrm{s}}\right) \mathrm{p}_{\hat{\gamma}_{\mathrm{s}}\,|\,\gamma}(\hat{\gamma}_{\mathrm{s}} \mid \gamma, N)\,\mathrm{d}\hat{\gamma}_{\mathrm{s}}. \tag{2.47}$$

by

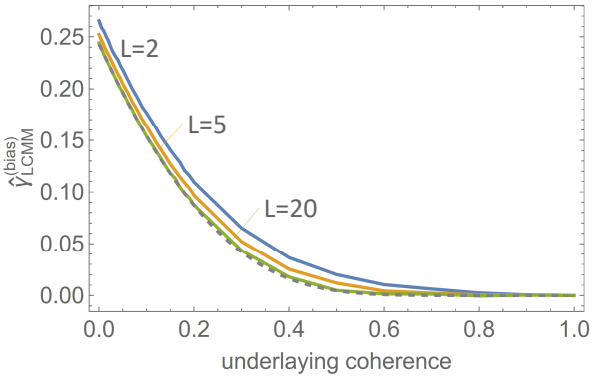
$$\mathrm{E}\{\hat{\gamma}_{\mathrm{LCMM}} \mid \gamma, N\} = \exp\left(\underline{m}_1\right). \tag{2.48}$$

Abdelfattah & Nicolas (2006) propose a numerical solution for (2.47). Nowadays, Mathematica can evaluate a closed-form solution for small N. The example for $N = 3$ is

$$\underline{m}_1 = -\frac{\gamma^4}{4} + \gamma^2 - \frac{3}{4}, \tag{2.49}$$



(a) Theoretical bias depending on the sample size N.



(b) Bias depending on the sample size L from simulations with $N = 10$.

Figure 2.8: Bias characteristic of LCMM estimator.

and for $N = 10$ it is[1]

$$\underline{m}_1 = \frac{\gamma^{18}}{18} - \frac{9\gamma^{16}}{16} + \frac{18\gamma^{14}}{7} - 7\gamma^{12} + \frac{63\gamma^{10}}{5} - \frac{63\gamma^8}{4} + 14\gamma^6 - 9\gamma^4 + \frac{9\gamma^2}{2} - \frac{7129}{5040}. \quad (2.50)$$

Using these results and (2.48), Fig. 2.8a shows that the bias is very large for small N. Also, Fig. 2.8b visualizes that many independent values, e.g., $L \geq 20$, need to be averaged in order to achieve the theoretical bias, which is plotted as a dashed line. In this example, the total number of i.i.d. samples $T = NL$ is 20 for $L = 2$, 50 for $L = 5$ and 200 for $L = 20$.

Zebker & Chen (2005) have developed a method to correct the bias fitting a polynomial to correlation estimates of simulated data as a function of true correlation and the number of looks in the estimate. This is a very practicable and powerful approach, as it represents the expected value of (2.15) favorably and corrects the bias.

More recently, Jiang et al. (2014) demonstrated the mitigation of the sample coherence bias with only several samples based on *Double Bootstrapping*. Basically, *Bootstrapping* is a non-parametric approach to statistical inference. It is powerful to estimate the bias of an estimate and a second iteration is proposed in order to estimate and mitigate the bias in the bias estimate. Generally, *Bootstrapping* is based on randomly sample N times with replacement from a given sample of size N. First, the original observations $\boldsymbol{X} = \{x_{1,i=1}, x_{2,i=1}, \ldots, x_{1,i=N}, x_{2,i=N}\}$ are R times sampled respectively resulting in the samples $\boldsymbol{X}_r^*$. Each $\boldsymbol{X}_r^*$ provides a sample estimate $\hat{\gamma}_r^*$ using (2.11). Second, M times random samples $\boldsymbol{X}_{r,m}^{**}$ from the first random samples $\boldsymbol{X}_r^*$ are taken and provide the sample coherence estimates $\hat{\gamma}_{r,m}^{**}$. The original sample $\boldsymbol{X}$ provides $\hat{\gamma}_{\mathrm{s}}$. Jiang et al. (2014) derived the estimation equation for the *Double Bootstrapping* estimate $\hat{\gamma}_{\mathrm{DB}}$

$$\hat{\gamma}_{\mathrm{DB}} = 3\hat{\gamma}_{\mathrm{s}} - \frac{3}{R}\sum_{r=1}^{R}\hat{\gamma}_r^* + \frac{1}{RM}\sum_{r=1}^{R}\sum_{m=1}^{M}\hat{\gamma}_{r,m}^{**}. \quad (2.51)$$

[1]Since the domain of γ is between 0 and 1, the large powers are also in this range.

This method is computationally demanding and requires $R(1+M)+1$ sample coherence estimations (2.11) of size N, where Jiang et al. (2014) suggests $R \geq 500$. Furthermore, the double bias correction introduces extra estimation variability which can be observed by a high estimator RMSE which is evident in column 2 of Table I reported by Adam (2022).

ML is an active field of research in its own right. The application of ML in remote sensing is comprehensively reviewed by Zhu et al. (2017) and Schmitt et al. (2023). In the context of SAR signal processing, Sica et al. (2021, 2023) apply Deep Learning and self-supervised learning for InSAR phase and coherence estimation with focus on phase denoising.

3 Methods for Coherence Magnitude Estimation

Starting point are random CCG processes $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ with an underlying coherence magnitude γ and InSAR phase φ represented by specific realizations $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. The specific realizations $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are available by coregistered samples of a homogenous area with N interferometric data pairs $\{\boldsymbol{x}_1 \mid x_{1,1}, \ldots, x_{1,N}\}$ and $\{\boldsymbol{x}_2 \mid x_{2,1}, \ldots, x_{2,N}\}$. Practically, the coherence estimators are allowed to use only the N samples, which do not fully represent the processes $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. However, some relevant parameters can be estimated from these samples by the respective sample estimators, e.g., the expected intensity of the channel with index k

$$\begin{aligned} \hat{I}_k &= \mathrm{E}\{|\boldsymbol{x}_k|^2\} \\ &= \frac{1}{N} \sum_{i=1}^{N} |x_{k,i}|^2 \end{aligned} \tag{3.1}$$

and the expected InSAR phase

$$\begin{aligned} \hat{\varphi}_\mathrm{s} &= \arg\left(\mathrm{E}\{\boldsymbol{x}_1 \boldsymbol{x}_2^*\}\right) \\ &= \arg\left(\sum_{i=1}^{N} x_{1,i}\, x_{2,i}^*\right). \end{aligned} \tag{3.2}$$

The data are assumed to be stationary, i.e., all sample pairs have one and the same joint probability distribution (2.22) with equal conditional parameters γ, φ, I_1 and I_2 and are independent. Such samples are named i.i.d.. Practically, the samples are free of phase components other than the underlying InSAR phase φ which is disturbed by independent noise. According to Bamler & Hartl (1998), the latter is characterized by the SNR

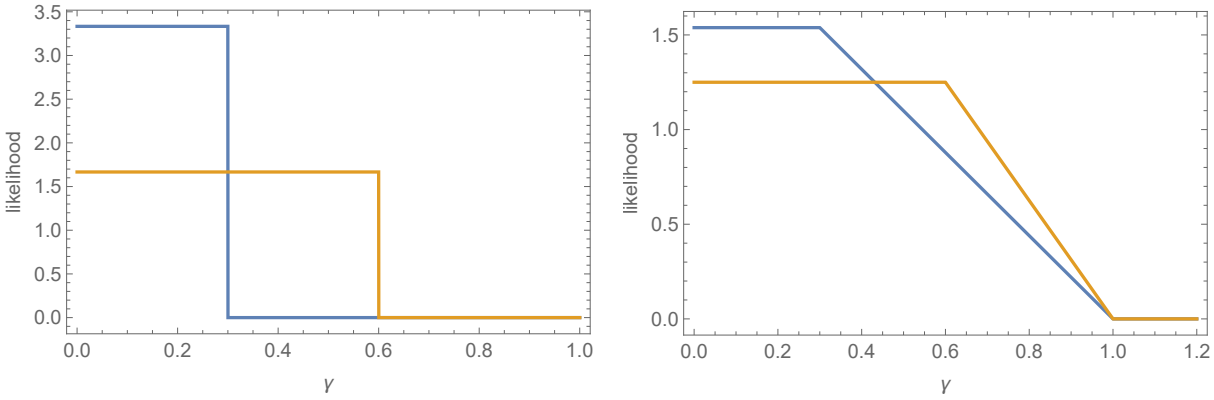
$$\gamma = \frac{1}{\sqrt{1 + \frac{1}{SNR_1}}\sqrt{1 + \frac{1}{SNR_2}}}. \tag{3.3}$$

As a matter of fact, the i.i.d. principle is not fulfilled by any available real SAR scene, regardless of the sensor. All focusing algorithms introduce a spectral weighting for sidelobe reduction and improving the signal to azimuth ambiguity ratio (SAAR) as described by Breit et al. (2008), Bourbigot (2016) in section A1.3.6 and Piantanida (2022) in section 9.7. Additionally, an oversampling is introduced in order to respect the fundamental Nyquist–Shannon sampling theorem and to simplify the InSAR signal processing, e.g., the interpolation. The latter effect is visualized by Hanssen & Bamler (1999) in Figs. 1 and

3. Hanssen & Bamler (1999) also describe the relevant parameters, i.e., the oversampling ratios in range r_{rg} and azimuth r_{az}, the range sampling frequency (RSF), the range bandwidth (RBW), the azimuth bandwidth (ABW) and the pulse-repetition frequency (PRF). These are related to each other by $r_{\mathrm{rg}} = {}^{RSF}/_{RBW}$ and $r_{\mathrm{az}} = {}^{PRF}/_{ABW}$. In practice, the number of SAR scene samples N_{pixel} and the number of effectively available i.i.d. samples N are related by $N = {}^{N_{\mathrm{pixel}}}/_{(r_{\mathrm{rg}} r_{\mathrm{az}})}$. Oversampling and spectral weighting cause autocorrelation in the channels and need to be reversed before coherence is estimated. This does not have to be done individually per statistically homogeneous area, but is better calculated only once for each SAR scene. Such preprocessing from SLC into i.i.d. SAR data is not a disadvantage of the developed techniques in particular. It is also necessary for all other coherence magnitude estimation methods. In fact, all known estimators work with i.i.d. samples, where independence implies zero autocorrelation of samples within the primary and secondary channel. In case of autocorrelation, the spatial arrangement of the samples (for InSAR on the 2D grid) would have to be considered by the estimators. To illustrate typical effective number of looks, Sentinel-1 acquired with Interferometric Wide swath mode beam IW2 is chosen as an example. An area of 5 azimuth times 4 range samples, i.e., $N_{\mathrm{pixel}} = 20$, corresponds to $N = 9$ and 6 azimuth times 7 range samples, i.e., $N_{\mathrm{pixel}} = 42$, reduce to $N = 20$ independent samples.

Although extensive research has been carried out on coherence estimation described in section 2.2 no single study exists which includes prior information on the coherence magnitude. All estimation methods developed in this thesis allow to include prior information. During development, attention was paid to supporting typical InSAR use cases and a straight forward parameterization of the prior. Finally, two prior types result from typical InSAR scenarios namely the strict and the less strict prior.

One example for the strict prior is omnipresent and results from the technical SAR system limit caused by thermal, quantization and SAR processing noise. Each operational SAR system characterizes this system limit by the Noise Equivalent Sigma Zero (NESZ) or the Noise Equivalent Beta Naught (NEBN). Both parameters describe the SNR_k of the



(a) Strict prior. *(b)* Less strict prior.

Figure 3.1: Distribution of priors $\mathrm{p}_\gamma(\gamma)$ with $\gamma_{max} = 0.3$ (blue) and $\gamma_{max} = 0.6$ (orange).

respective data samples in each channel k

$$\begin{aligned} SNR_k &= \frac{\sigma_k^0}{NESZ} \\ &= \frac{\beta_k^0}{NEBN}. \end{aligned} \tag{3.4}$$

The respective channel's radar brightness (Beta Naught) $\beta_k^0 = \hat{I}_k$ or the respective backscatter coefficient (Sigma Naught) $\sigma_k^0 = \hat{I}_k \sin(\theta_0)$ are estimated by the expected channel intensities (3.1) and the corresponding substitution in (3.4) provides the best possible SNR_1 and SNR_2. Equation (3.3) transforms both SNRs into the maximum coherence $\gamma_{\max}$ which can be received interferometrically. This prior can be modeled by a uniform distribution $\mathrm{U}(0, \gamma_{\max})$

$$\mathrm{p}_\gamma(\gamma) = \mathrm{U}(0, \gamma_{\max}) = \begin{cases} \frac{1}{\gamma_{\max}} & 0 \le \gamma \le \gamma_{\max} \\ 0 & \text{otherwise} \end{cases}. \tag{3.5}$$

Another prior information is available in InSAR based on stacks of interferograms. For example, $\gamma_{\max}$ can be estimated from an initial coherence matrix (Ferretti et al. 2011, Fig. 5) which can straight forward be converted into the best possible coherence as a function of acquisition time difference. Depending on the $\gamma_{\max}$ accuracy and the likelihood that the underlying coherence is above $\gamma_{\max}$, the less strict or strict prior should be selected. In this thesis, the less strict prior is defined by

$$\mathrm{p}_\gamma(\gamma) = \begin{cases} \frac{2}{(1+\gamma_{\max})} & 0 \le \gamma \le \gamma_{\max} \\ \frac{2}{1-\gamma_{\max}^2} - \frac{2\gamma}{1-\gamma_{\max}^2} & \gamma_{\max} < \gamma \le 1 \\ 0 & \text{otherwise} \end{cases}. \tag{3.6}$$

Figure 3.1a shows the distribution of the strict and Fig. 3.1b of the less strict prior. The strict prior limits the estimates inside the assumed range and the less strict prior favors estimates in this range. In principle, any shape of prior can easily be implemented in the developed estimators. The uninformative prior is the strict prior with $\gamma_{\max} = 1$.

3.1 Empirical Bayesian Estimation

The method described in this section was originally published by the author in (Adam 2022).

The Bayesian analysis (Berger 1985) is based on the *posterior* distribution of the processes underlying true coherence γ which is constructed by Bayes' theorem from the data samples D

$$\mathrm{p_{AP}}(\gamma \mid D) = \frac{\mathrm{p_L}(D \mid \gamma)\ \mathrm{p_P}(\gamma)}{\int_\gamma \mathrm{p_L}(D \mid \gamma)\ \mathrm{p_P}(\gamma)\,\mathrm{d}\gamma}. \tag{3.7}$$

In InSAR, the data samples are available in pairs from the two SLC scenes $D = \{x_{1,i}, x_{2,i} \mid i = 1, \ldots, N\}$. That is, D denotes all pairs of samples of a homogeneous region. The likelihood function $\mathrm{p_L}(D \mid \gamma)$ is proportional to the probability of receiving the data D if the coherence γ was the basis. $\mathrm{p_P}(\gamma)$ is the prior and corresponds to a hypothesis on the coherence γ. In the equation above, the denominator is named evidence. It is the normalization which makes the *posterior* a proper probability density distribution. It is apparent that the *posterior* coherence γ is a random variable described by the *posterior* $\mathrm{p_{AP}}(\gamma \mid D)$ in contrast to the γ in the likelihood $\mathrm{p_L}(D \mid \gamma)$. This is the reason why the coherence is estimated to be the mode for the maximum *a posteriori* (MAP) estimator

$$\hat{\gamma}_{\mathrm{MAP}} = \operatorname*{argmax}_\gamma \mathrm{p_{AP}}(\gamma \mid D) \tag{3.8}$$

or the expected value of the *posterior* distribution for the expected *a posteriori* (EAP) estimator

$$\hat{\gamma}_{\mathrm{EAP}} = \mathrm{E}\{\mathrm{p_{AP}}(\gamma \mid D)\} \tag{3.9}$$

or the median value of the *posterior* distribution for the median *a posteriori* (MEDAP) estimator

$$\hat{\gamma}_{\mathrm{MEDAP}} = \mathrm{Median}\{\mathrm{p_{AP}}(\gamma \mid D)\}. \tag{3.10}$$

The mode, mean and median of the *posterior* distribution $\mathrm{p_{AP}}(\gamma \mid D)$ make a difference because of the skewness of this distribution. In order to evaluate $\mathrm{p_{AP}}(\gamma \mid D)$, the right hand side terms of (3.7), i.e., likelihood, prior and evidence, need to be computed.

Assuming N i.i.d. random variables $x_{1,i}$ and $x_{2,i}$ in a homogeneous area of each SAR scene with pixel index i, the likelihood function $\mathrm{p_L}(D \mid \gamma)$ in (3.7) is the product of the N individually joint PDFs $\mathrm{p}(x_{1,i}, x_{2,i} \mid \gamma, \varphi)$

$$\mathrm{p_L}(D \mid \gamma) = \prod_{i=1}^{N} \mathrm{p}(x_{1,i}, x_{2,i} \mid \gamma, \varphi). \tag{3.11}$$

The specific joint PDFs are given by (2.22) as a function of the sample pair's amplitudes $a_{k,i} = |x_{k,i}|$ and of the pixel-wise interferometric phase difference $\varphi_i = \arg(x_{2,i}) - \arg(x_{1,i}) = \phi_{2,i} - \phi_{1,i}$ given the underlying coherence γ, the true interferometric phase φ and the expected (true) channel intensities I_1 and I_2, i.e.,

$$\mathrm{p_L}(D \mid \gamma) = \prod_{i=1}^{N} \mathrm{p_{InSAR_1}}(a_{1,i}, a_{2,i}, \varphi_i \mid \gamma, \varphi, I_1, I_2). \tag{3.12}$$

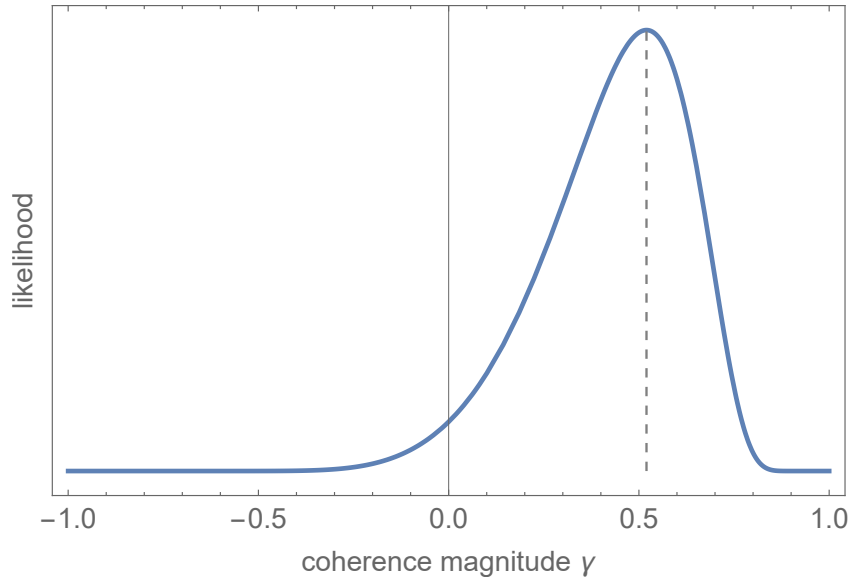


Figure 3.2: Example likelihood (3.12) with $\hat{\gamma}_\mathrm{s} = 0.52$ for $N = 5$ and $\gamma = 0.4$.

As an example, Fig. 3.2 visualizes a likelihood function for five simulated CCG process sample pairs with true $\gamma = 0.4$. The MLE overestimates the underlying simulated coherence and is in this particular example $\hat{\gamma}_\mathrm{s} = 0.52$. It is noticeable in plot 3.2 that negative coherences also have a likelihood, although negative values are not meaningful according to the definition (2.10). It is clear, the true coherence magnitude γ and the MLE argmax(.) are always in the domain $\{\gamma \mid 0 \leq \gamma \leq 1\}$. However, transforming the likelihood by using the prior and the evidence requires to expand the domain of the coherence parameter. All previous studies have not dealt with this empirical generalization. The expanded domain of coherence and the way in which it is carried out leads to the name *empirical* Bayesian estimation.

In (3.7), the prior PDF $\mathrm{p}_\mathrm{P}(\gamma)$ is a key element. As far as the conventional Bayesian framework is concerned, it is independent of the observed data. This thesis develops an approach where the prior distribution is estimated from the data, i.e., the sample estimator (2.11). It is another reason for the name *empirical* Bayes estimation.

Subject is to develop the prior $\mathrm{p}_\mathrm{P}(\gamma)$ in (3.7) which depends on the sample estimator outcome $\hat{\gamma}_s$. For this reason, it is a conditional probability distribution $\mathrm{p}_\mathrm{P}(\gamma) = \mathrm{p}_{\gamma\,|\,\hat{\gamma}_\mathrm{s}}(\gamma \mid \hat{\gamma}_\mathrm{s}, N)$ expressing the underlying coherence γ probability given the sample estimate $\hat{\gamma}_s$ based on N sample pairs. It needs to include the available bias information. For the sample coherence estimator (2.11), Carter et al. (1973) and Touzi & Lopes (1996) deduced the respective conditional probability distribution for $N > 2$ samples provided by (2.12). The product rule relates the joint PDF $\mathrm{p}_{\hat{\gamma}_\mathrm{s},\gamma}(\hat{\gamma}_\mathrm{s}, \gamma \mid N)$ of the sample coherence $\hat{\gamma}_\mathrm{s}$ and underlying coherence γ with (2.12)

$$\mathrm{p}_{\hat{\gamma}_\mathrm{s},\gamma}(\hat{\gamma}_\mathrm{s}, \gamma \mid N) = \mathrm{p}_{\hat{\gamma}_\mathrm{s}\,|\,\gamma}(\hat{\gamma}_\mathrm{s} \mid \gamma, N)\, \mathrm{p}_\gamma(\gamma \mid N). \tag{3.13}$$

The term $\mathrm{p}_\gamma(\gamma \mid N)$ is the data independent prior and is simplified into $\mathrm{p}_\gamma(\gamma)$ due to the independence of γ and N, i.e., N has no effect on the true coherence γ and vice versa.

First, no assumption is made about the underlying coherence and this prior is named uninformative prior. Of course, a general hypothesis on the maximum coherence magnitude γ_{max} can be made also. This information corresponds to a data independent prior and is established by a modification of the uninformative prior Bayes approach and is introduced subsequently.

3.1.1 Uninformative Prior

For a generally applicable estimator, a uniform distribution $\mathrm{U}(-1,1)$, i.e., a flat prior is selected for $\mathrm{p}_\gamma(\gamma)$

$$\mathrm{p}_\gamma(\gamma) = \mathrm{U}(-1,1) = \begin{cases} 1/2 & -1 \leq \gamma \leq 1 \\ 0 & \text{otherwise} \end{cases}. \tag{3.14}$$

In doing so, the joint PDF $\mathrm{p}_{\hat{\gamma}_\mathrm{s},\gamma}(\hat{\gamma}_\mathrm{s},\gamma \mid N)$ of $\hat{\gamma}_\mathrm{s}$ and γ is proportional to the conditional probability distribution of the estimated coherence magnitude $\hat{\gamma}_\mathrm{s}$ in (2.12)

$$\mathrm{p}_{\hat{\gamma}_\mathrm{s},\gamma}(\hat{\gamma}_\mathrm{s},\gamma \mid N) = \frac{1}{2}\,\mathrm{p}_{\hat{\gamma}_\mathrm{s}\mid\gamma}(\hat{\gamma}_\mathrm{s} \mid \gamma, N). \tag{3.15}$$

An example joint PDF of $\hat{\gamma}_\mathrm{s}$ and γ is visualized in Fig. 3.3

From the joint PDF $\mathrm{p}_{\hat{\gamma}_\mathrm{s},\gamma}(\hat{\gamma}_\mathrm{s},\gamma \mid N)$, the conditional probability distribution $\mathrm{p}_{\gamma\mid\hat{\gamma}_\mathrm{s}}(\gamma \mid \hat{\gamma}_\mathrm{s}, N)$ providing the empirical prior $\mathrm{p}_\mathrm{P}(\gamma)$ can be derived with Bayes' theorem

$$\mathrm{p}_\mathrm{P}(\gamma) = \mathrm{p}_{\gamma\mid\hat{\gamma}_\mathrm{s}}(\gamma \mid \hat{\gamma}_\mathrm{s}, N) = \frac{\mathrm{p}_{\hat{\gamma}_\mathrm{s},\gamma}(\hat{\gamma}_\mathrm{s},\gamma \mid N)}{\mathrm{p}_{\hat{\gamma}_\mathrm{s}}(\hat{\gamma}_\mathrm{s} \mid N)}. \tag{3.16}$$

In the equation above, it is difficult to make assumptions about the term $\mathrm{p}_{\hat{\gamma}_\mathrm{s}}(\hat{\gamma}_\mathrm{s} \mid N)$. It is the PDF of observing $\hat{\gamma}_\mathrm{s}$ from N interferometric samples. Practically, for *a posteriori*

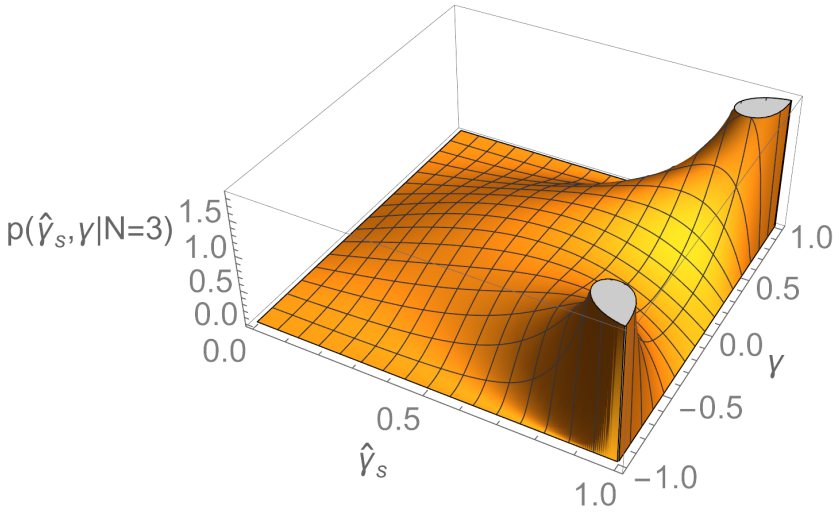


Figure 3.3: Example joint PDF of $\hat{\gamma}_\mathrm{s}$ and γ, i.e., $\mathrm{p}_{\hat{\gamma}_\mathrm{s},\gamma}(\hat{\gamma}_\mathrm{s},\gamma \mid N = 3)$.

based estimators, we are interested in a prior PDF $\mathrm{p_P}(\gamma)$ which integrates to 1 whereas $\mathrm{p}_{\gamma|\hat{\gamma}_\mathrm{s}}(\gamma \mid \hat{\gamma}_\mathrm{s}, N)$ typically has not this characteristic. The following marginalization of $\mathrm{p}_{\hat{\gamma}_\mathrm{s},\gamma}(\hat{\gamma}_\mathrm{s}, \gamma \mid N)$ with respect to γ makes the expression $\mathrm{p_P}(\gamma)$ a valid PDF and comparison with (3.7) provides the PDF of observing $\hat{\gamma}_\mathrm{s}$ given the number of samples

$$\begin{aligned}\mathrm{p}_{\hat{\gamma}_\mathrm{s}}(\hat{\gamma}_\mathrm{s} \mid N) &= \int_{-1}^{1} \mathrm{p}_{\hat{\gamma}_\mathrm{s},\gamma}(\hat{\gamma}_\mathrm{s}, \gamma \mid N) d\gamma \\ &= \frac{\hat{\gamma}_\mathrm{s}(N-1)(1-\hat{\gamma}_\mathrm{s}^2)^{N-2}\sqrt{\pi}\,N!}{\Gamma\left(N+\frac{3}{2}\right)} {}_3\mathrm{F}_2\left(\frac{1}{2}, N, N; 1, N+\frac{3}{2}; \hat{\gamma}_\mathrm{s}^2\right).\end{aligned} \tag{3.17}$$

$\mathrm{p}_{\hat{\gamma}_\mathrm{s}}(\hat{\gamma}_\mathrm{s} \mid N)$ is plotted in Fig. 3.4 and the graphs correspond with the known characteristic of the sample coherence estimator. For N very small, e.g., $N = 3$, the likelihood for large $\hat{\gamma}_\mathrm{s}$ is higher than for each smaller $\hat{\gamma}_\mathrm{s}$ because all sample estimates are biased towards higher coherence values. For $N = 9$, the plot shows a similar likelihood for sample estimates above 0.4. For N very large, e.g., $N = 100$, the likelihood above 0.3 is similar. In fact, all graphs visualize the effect that small coherences are unlikely to be estimated due to the bias. The plot also shows, the larger the sample size, the smaller the coherence that can be reliably estimated.

From (2.12) and (3.17), the empirical prior results

$$\begin{aligned}\mathrm{p_P}(\gamma) &= \mathrm{p}_{\gamma|\hat{\gamma}_\mathrm{s}}(\gamma \mid \hat{\gamma}_\mathrm{s}, N) \\ &= \frac{(1-\gamma^2)^N\, {}_2\mathrm{F}_1(N, N; 1; \gamma^2\,\hat{\gamma}_\mathrm{s}^2)\;\Gamma\left(N+\frac{3}{2}\right)}{{}_3\mathrm{F}_2\left(\frac{1}{2}, N, N; 1, N+\frac{3}{2}; \hat{\gamma}_\mathrm{s}^2\right)\;\sqrt{\pi}\,N!}.\end{aligned} \tag{3.18}$$

The data dependent prior is in principle a bimodal PDF which becomes unimodal for small $\hat{\gamma}_\mathrm{s}$ and is visualized in Fig. 3.5. With (3.11) and (3.18), the *posterior* distribution

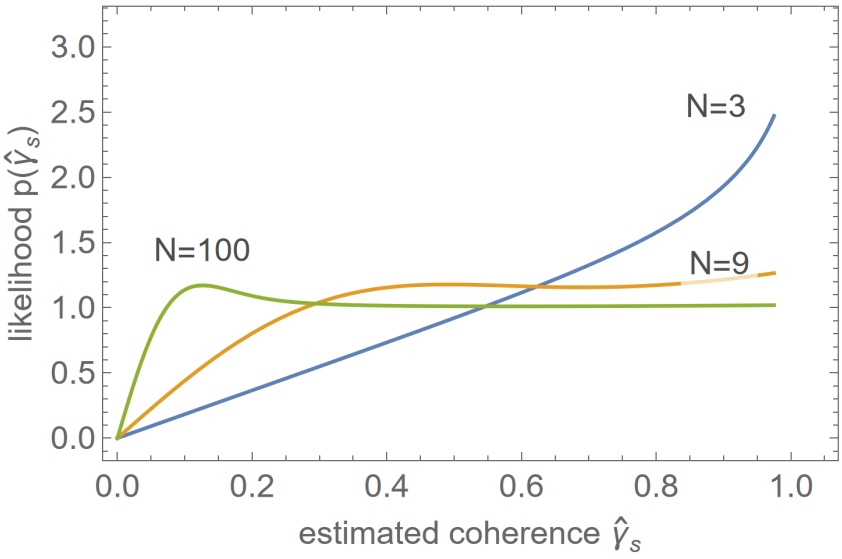
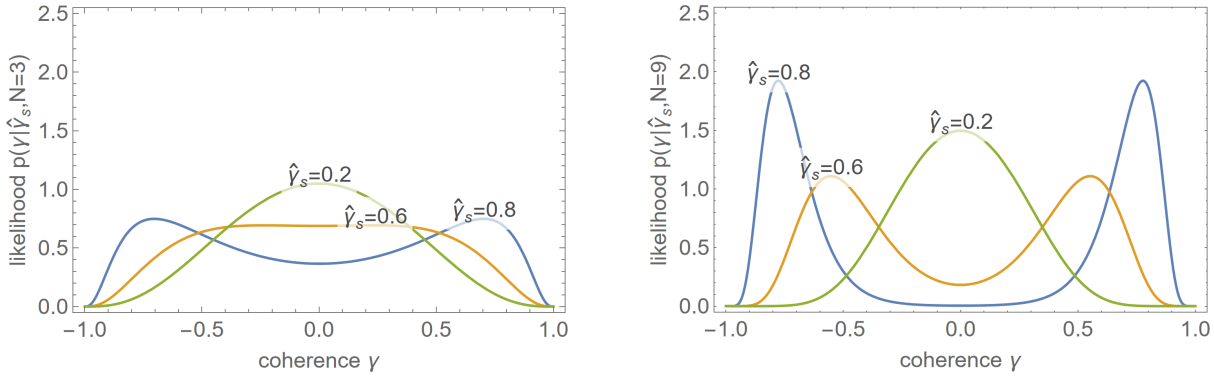


Figure 3.4: PDF of observing the sample estimate $\hat{\gamma}_\mathrm{s}$.



(a) $N = 3$. *(b)* $N = 9$.

Figure 3.5: Example empirical priors based on an uninformative general prior.

(3.7) is practically evaluated by

$$\mathrm{p_{AP}}(\gamma \mid a_{1,i=1}, a_{2,i=1}, \ldots, a_{1,i=N}, a_{2,i=N}, \varphi_{i=1} \ldots, \varphi_{i=N}) = \frac{1}{m} \frac{(1-\gamma^2)^N \, {}_2\mathrm{F}_1(N, N; 1; \gamma^2 \hat{\gamma}_\mathrm{s}^2) \; \Gamma\left(N + \frac{3}{2}\right)}{{}_3\mathrm{F}_2\left(\frac{1}{2}, N, N; 1, N + \frac{3}{2}; \hat{\gamma}_\mathrm{s}^2\right) \sqrt{\pi}\, N!} \prod_{i=1}^{N} \mathrm{p_{InSAR_1}}(a_{1,i}, a_{2,i}, \varphi_i \mid \gamma, I_1, I_2, \hat{\varphi}_\mathrm{s}). \tag{3.19}$$

The evidence, i.e., the marginal likelihood $m = \int_{-1}^{1} \mathrm{p_L}(.)\, \mathrm{p_P}(.)\, d\gamma$ is not needed for the coherence estimation by the MAP and MEDAP principle. However, for the EAP method, it is straight forward implemented within the expected value calculation. It is only needed once for the calculation of this value and a numerical integration works without any problems.

3.1.2 Informative Priors

The principle of the estimator with uninformative prior is extended. Two general priors are designed to express practically available information. In order to be used in the empirical estimation, (3.5) and (3.6) need to be adjusted. First, a strict maximum underlying coherence γ_max is assumed, e.g., by knowing the best possible primary and secondary SNRs from the relation (3.3). The respective data independent prior is

$$\mathrm{p}_\gamma(\gamma) = \mathrm{U}(-\gamma_\mathrm{max}, \gamma_\mathrm{max}) = \begin{cases} 1/(2\gamma_\mathrm{max}) & -\gamma_\mathrm{max} \leq \gamma \leq \gamma_\mathrm{max} \\ 0 & \text{otherwise} \end{cases}. \tag{3.20}$$

Second, a less strict maximum underlying coherence γ_max is assumed, e.g., a maximum coherence is known from many measurements. For InSAR based on time series, this is a typical situation with data stacks covering more than a decade. A linearly decreasing

function outside of the assumed range makes the prior less strict

$$\mathrm{p}_\gamma(\gamma) = \begin{cases} 1/(1+\gamma_{\max}) & -\gamma_{\max} \leq \gamma \leq \gamma_{\max} \\ 1/(1-\gamma_{\max}^2) + \gamma/(1-\gamma_{\max}^2) & -1 \leq \gamma < -\gamma_{\max} \\ 1/(1-\gamma_{\max}^2) - \gamma/(1-\gamma_{\max}^2) & \gamma_{\max} < \gamma \leq 1 \\ 0 & \text{otherwise} \end{cases} . \tag{3.21}$$

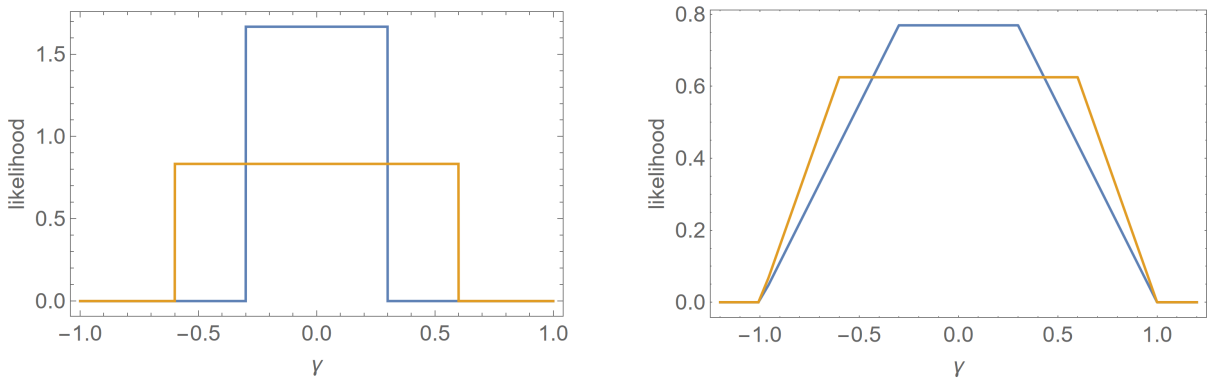
Both informative priors are visualized in Fig. 3.6.

In order to get the data dependent prior $\mathrm{p}_{\gamma\,|\,\hat{\gamma}_\mathrm{s}}(\gamma \mid \hat{\gamma}_\mathrm{s}, N)$ equivalent to (3.16), the joint probability $\mathrm{p}_{\hat{\gamma}_\mathrm{s},\gamma}(\hat{\gamma}_\mathrm{s}, \gamma \mid N) = \mathrm{p}_{\hat{\gamma}_\mathrm{s}\,|\,\gamma}(\hat{\gamma}_\mathrm{s} \mid \gamma, N)\,\mathrm{p}_\gamma(\gamma)$ needs to be evaluated. This is straight forward implemented with (2.12) and (3.20) or (3.21). However, the respective marginalization integral over γ for the normalization $\mathrm{p}_{\hat{\gamma}_\mathrm{s}}(\hat{\gamma}_\mathrm{s})$ cannot be solved directly and numerical integration in a later calculation step works well. It needs to be evaluated once per estimation and can be computed together with the overall normalization (i.e., evidence) which is finally needed for the *a posteriori* PDF. Analogous to (3.19) which describes the general situation, the *posterior* distribution for the more informative prior $\mathrm{p}_\gamma(\gamma)$ from (3.20) or (3.21) is

$$\begin{aligned} &\mathrm{p}_\mathrm{AP}(\gamma \mid a_{1,i=1}, a_{2,i=1}, \ldots, a_{1,i=N}, a_{2,i=N}, \varphi_{i=1} \ldots, \varphi_{i=N}) = \\ &\quad \frac{\mathrm{p}_{\hat{\gamma}_\mathrm{s}\,|\,\gamma}(\hat{\gamma}_\mathrm{s} \mid \gamma, N)\,\mathrm{p}_\gamma(\gamma)}{m} \prod_{i=1}^{N} \mathrm{p}_{\mathrm{InSAR}_1}(a_{1,i}, a_{2,i}, \varphi_i \mid \gamma, I_1, I_2, \hat{\varphi}_\mathrm{s}). \end{aligned} \tag{3.22}$$

In practice, the equations (3.19) and (3.22) can only be implemented directly without normalization m. To denote these non-normalized *posterior* functions, the symbol $\mathrm{q}_\mathrm{AP}(.)$ is used below

$$\begin{aligned} &\mathrm{q}_\mathrm{AP}(\gamma \mid a_{1,i=1}, a_{2,i=1}, \ldots, a_{1,i=N}, a_{2,i=N}, \varphi_{i=1} \ldots, \varphi_{i=N}) = \\ &\quad \mathrm{p}_{\hat{\gamma}_\mathrm{s}\,|\,\gamma}(\hat{\gamma}_\mathrm{s} \mid \gamma, N)\,\mathrm{p}_\gamma(\gamma) \prod_{i=1}^{N} \mathrm{p}_{\mathrm{InSAR}_1}(a_{1,i}, a_{2,i}, \varphi_i \mid \gamma, I_1, I_2, \hat{\varphi}_\mathrm{s}). \end{aligned} \tag{3.23}$$



(a) Strict prior. *(b)* Less strict prior.

Figure 3.6: Data independent informative prior $\gamma_{\max} = 0.3$ (blue) and $\gamma_{\max} = 0.6$ (orange).

Effectively, the evidence i.e., the marginal likelihood is computed from the non-normalized *posterior* functions

$$m = \int_{-1}^{1} \mathrm{q}_{\mathrm{AP}}(.)\, d\gamma \tag{3.24}$$

and a numerical integration is fine.

3.1.3 Estimation of Coherence Magnitude

For the coherence estimation, the statistic of the *a posteriori* PDF $\mathrm{p}_{\mathrm{AP}}(.)$ (3.19) or (3.22) is analyzed based on the non-normalized versions $\mathrm{q}_{\mathrm{AP}}(.)$.

MAP estimation The MAP estimation (3.8) is implemented by a numerical 1-D maximization of $\mathrm{q}_{\mathrm{AP}}(.)$ with the constraint $0 \leq \hat{\gamma}_{\mathrm{MAP}} \leq \gamma_{\max}$. For the uninformative and the less strict prior, the constraint is $\gamma_{\max} = 1$ and for the strict prior, it is the specific $\gamma_{\max}$ parameter

$$\hat{\gamma}_{\mathrm{MAP}} = \underset{\gamma}{\operatorname{argmax}}\, \mathrm{q}_{\mathrm{AP}}(\gamma \mid D) \tag{3.25}$$

The respective implementation is straight forward as the *posterior* function is unimodal in the range $0 - 1$.

EAP estimation The EAP estimation (3.9) is implemented by numerical integration

$$\hat{\gamma}_{\mathrm{EAP}} = \frac{1}{m} \int_{-1}^{1} \gamma\, \mathrm{q}_{\mathrm{AP}}(.)\, d\gamma \tag{3.26}$$

with the evidence m calculated independently from (3.24). The specific estimators $\hat{\gamma}_{\mathrm{EAP}}$, $\hat{\gamma}_{\mathrm{EAPLSP}_{\gamma_{\max}}}$ and $\hat{\gamma}_{\mathrm{EAPSP}_{\gamma_{\max}}}$ use directly equation (3.19), or (3.23) with (3.21) or (3.20) respectively.

MEDAP estimation The MEDAP estimation (3.10) is implemented by bisection on the left and right areas separated by $\hat{\gamma}_{\mathrm{MEDAP}}$ under the *posterior* functions. This is the reason why the marginal m is not relevant for the median computation

$$\int_{-1}^{\hat{\gamma}_{\mathrm{MEDAP}}} \mathrm{q}_{\mathrm{AP}}(.)\, d\gamma \stackrel{!}{=} \int_{\hat{\gamma}_{\mathrm{MEDAP}}}^{1} \mathrm{q}_{\mathrm{AP}}(.)\, d\gamma. \tag{3.27}$$

The $\stackrel{!}{=}$ indicates the condition that must be fulfilled. However, this straight forward implementation requires two integrations per bisection step. Using (3.24), the bisection can be simplified into

$$\int_{\hat{\gamma}_{\mathrm{MEDAP}}}^{1} \mathrm{q}_{\mathrm{AP}}(.)\, d\gamma \stackrel{!}{=} \frac{m}{2} \tag{3.28}$$

and requires only one integration per bisection step.

3.2 Machine Learning Estimation

The method described in this section was originally published by the author in (Adam 2023).

ML can solve many tasks, and there are numerous specialized areas of ML. A common classification is that of supervised, unsupervised, and reinforcement learning. To implement coherence estimation, supervised learning must be selected from the ML subdomains because it learns the objective function from labeled datasets. It is therefore suitable for classification or regression and can be implemented practically, e.g., as neural network (NN) or decision tree based. Zhu et al. (2017) describe some NN architectures, e.g., the Convolutional Neural Network (CNN), the Recurrent Neural Network (RNN) and the Generative Adversarial Network (GAN). Tree-based methods use decision trees where each node is a test on an attribute, each branch is an outcome of the test and the leaf of a node is the regression or classification outcome. Ensemble methods combine many decision trees and improve the performance and avoid overfitting. Corresponding ML techniques are Random Forrest and Gradient Boosting. A Random Forrest (Ho 1995) builds many decision trees from different subsets of the data and averages the outcomes. Gradient Boosting (Freund & Schapire 1996) builds many week decision trees in a clever sequence, correcting the errors of the previous decision trees. eXtreme Gradient Boosting (XGBoost) (Chen & Guestrin 2016) and Light Gradient-Boosting Machine (LightGBM) (Meng et al. 2016) are popular implementations of gradient boosting.

Subject is to develop a new method for estimation of the coherence magnitude $\hat{\gamma}$ based on the random CCG processes $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ with specific realizations $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$

$$\begin{aligned}\hat{\gamma} =& f(\boldsymbol{x}_1, \boldsymbol{x}_2) \\ =& f(x_{1,i=1}, x_{2,i=1}, \ldots, x_{1,i=N}, x_{2,i=N}).\end{aligned} \tag{3.29}$$

The ML approach results in a non-parametric method. Because input samples are mapped to a continuous output value, this ML task corresponds to the regression problem in contrast to classification. Figure 3.7 visualizes the basic principle and components of the development. Initially, non-parametric estimators $\hat{f}_N(x_{1,i=1}, x_{2,i=1}, \ldots, x_{1,i=N}, x_{2,i=N})$ are automatically generated from simulations and by supervised learning for any practically occurring number of samples N. In the operational system, this estimator $\hat{f}_N(.)$ is then used with N single look i.i.d. interferometric samples. Both system components are described in the following.

3.2.1 Generation of Estimators

For all required sample sizes N, ML provides a representative non-parametric model $\hat{f}_N(.)$. That means, there is no assumption about the function shape and the internal dependencies of the extracted features. As a result, a previously unknown number of internal parameters is required to represent the model and, accordingly, a lot of training data and computational effort are necessary for the learning. However, this does not pose

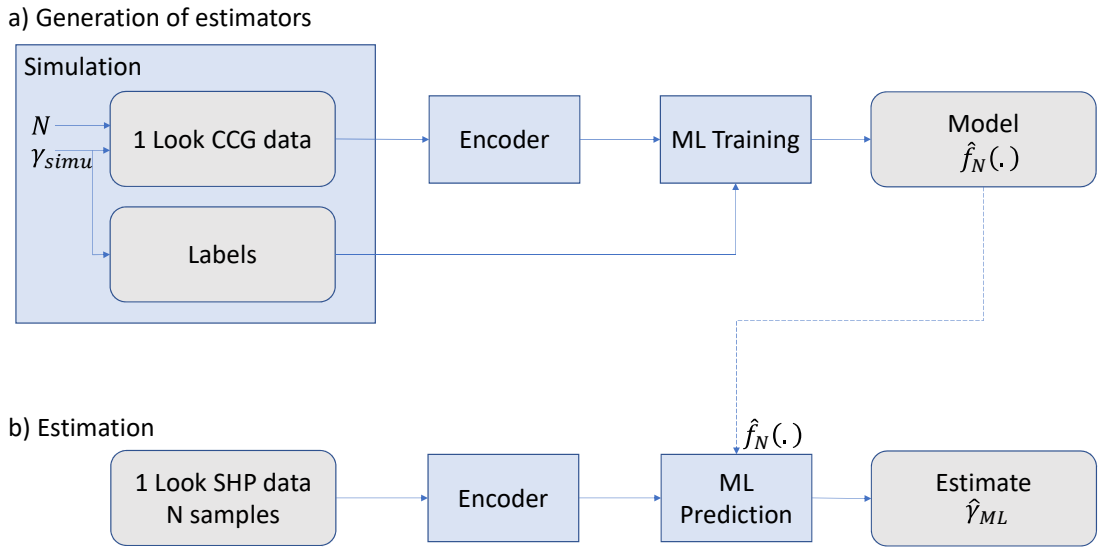


Figure 3.7: Principle of the ML framework for coherence magnitude estimation.

a problem, since the corresponding data can be simulated in practically any quantity and the theoretically infinite number of possible variants of input data can be restricted in terms of quantity.

Simulation

As pointed out by Goodman (1976) and Just & Bamler (1994), we can limit ourselves to CCG signals for medium resolution SAR. Starting point is the 2×2 covariance matrix $\mathbf{\Sigma}$ which describes the relation of the respective CCG processes $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$

$$\mathbf{\Sigma} = \mathrm{Cov}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \begin{pmatrix} (a_1)^2 & a_1\, a_2\, \gamma\, \mathrm{e}^{\mathrm{j}\varphi} \\ a_1\, a_2\, \gamma\, \mathrm{e}^{-\mathrm{j}\varphi} & (a_2)^2 \end{pmatrix}. \tag{3.30}$$

It is defined by the simulation parameters a_1, a_2, which are the CCG processes' expected amplitudes, and the complex coherence $\gamma e^{j\varphi}$. The coherence magnitude γ_{simu} is substituted for γ and is also used as the ML label for the respective simulated data. The matrix above contains the expected intensities on the diagonal and the covariances on the off-diagonal. φ is the true interferometric phase.

First of all, the square, positive definite and Hermitian covariance matrix $\mathbf{\Sigma}$ is decomposed

$$\mathbf{\Sigma} = \boldsymbol{A}\boldsymbol{A}^H. \tag{3.31}$$

The superscript H denotes the conjugate transpose of the complex matrix. Practically, this operation can be performed using singular value decomposition (SVD), Schur decomposition or Cholesky decomposition.

The SVD

$$\boldsymbol{U}\boldsymbol{W}\boldsymbol{V} = \mathrm{SVD}(\mathbf{\Sigma}) \tag{3.32}$$

results in $\boldsymbol{A} = \boldsymbol{U}\sqrt{\boldsymbol{W}}$, where $\sqrt{\boldsymbol{W}}$ denotes the element-by-element square root of the diagonal.

The Schur decomposition produces an orthonormal matrix $\boldsymbol{Q}$ and an upper triangular matrix $\boldsymbol{T}$

$$\boldsymbol{Q}\boldsymbol{T} = \text{Schur}(\boldsymbol{\Sigma}). \tag{3.33}$$

This gives $\boldsymbol{A} = \boldsymbol{Q}\sqrt{\boldsymbol{T}}$, where $\sqrt{\boldsymbol{T}}$ is the element-wise square root of all matrix entries in the complex domain.

With the Cholesky decomposition

$$\boldsymbol{L}\boldsymbol{L}^H = \text{Cholesky}_L(\boldsymbol{\Sigma}), \tag{3.34}$$

the product of a lower triangular $\boldsymbol{L}$ and its conjugate transpose matrix $\boldsymbol{L}^H$ arises. In this case $\boldsymbol{A} = \boldsymbol{L}$. In case, the library provides the upper triangular matrix $\boldsymbol{U}$ such that

$$\boldsymbol{U}^H\boldsymbol{U} = \text{Cholesky}_U(\boldsymbol{\Sigma}) \tag{3.35}$$

then $\boldsymbol{A} = \boldsymbol{U}^H$.

Next, a complex matrix $\boldsymbol{Z} \in \mathbb{C}^{2\times N}$ of independent CCG random variables

$$\begin{aligned} z_{k,i} &= \mathfrak{Re}(z_{k,i}) + j\mathfrak{Im}(z_{k,i}) \\ \mathfrak{Re}(z_{k,i}) &\sim \mathcal{N}(0, 1/\sqrt{2}) \\ \mathfrak{Im}(z_{k,i}) &\sim \mathcal{N}(0, 1/\sqrt{2}) \end{aligned} \tag{3.36}$$

is created. $\mathcal{N}(0, 1/\sqrt{2})$ denotes the normal distribution with zero mean and standard deviation $1/\sqrt{2}$. The simulated interferometric data pair corresponds to the complex matrix $\boldsymbol{S} \in \mathbb{C}^{2\times N}$ calculated by

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{x}_1^T \\ \boldsymbol{x}_2^T \end{pmatrix} = \begin{pmatrix} — & x_{1,i} & — \\ — & x_{2,i} & — \end{pmatrix} = \boldsymbol{A}\boldsymbol{Z}. \tag{3.37}$$

This principle of transforming the covariance matrix $\boldsymbol{\Sigma}$ into an interferometric data pair $\boldsymbol{S}$ can be applied to the simulation of InSAR data stacks. The dimensions of the covariance and the CCG matrix $\boldsymbol{Z}$ need to be increased accordingly. All three decompositions were implemented and finally the SVD was used for this thesis.

Encoder

The encoder transforms the input data and has two preprocessing functions: a) reduce redundancies and b) convert the input into an advantageous data representation.

Ideally, the signal entering the ML training includes all appropriate features and recognizable patterns, and is a data representation without ambiguity or redundancy. This makes the ML training algorithm more precise and computationally efficient, and requires less computer memory.

Redundancy Reduction: From the original simulation input parameters (3.30) $\{a_1 \mid 0 < a_1 < \infty\}$, $\{a_2 \mid 0 < a_2 < \infty\}$, $\{\gamma \mid 0 \leq \gamma \leq 1\}$ and $\{\varphi \mid -\pi < \varphi \leq \pi\}$ with their unrestricted domains and the unlimited possible combinations, we can see that there is an infinite number of possible input data sets. Generating and training all this is not realistic. As will be shown shortly, the expected channel amplitudes and the expected interferometric phase cause redundancies in the signal representation. Indeed the correlation coefficient $\rho_{\boldsymbol{X}_1,\boldsymbol{X}_2}$ is independent of change of origin, e.g., by real numbers b and d and scale of the data, e.g., by real numbers a and c

$$\rho_{X_1,X_2} = \mathrm{corr}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \mathrm{corr}(a\boldsymbol{X}_1 + b, c\boldsymbol{X}_2 + d). \tag{3.38}$$

This means that by scaling the amplitudes

$$x_{1,i} = \frac{x_{1,i}}{\max(|\boldsymbol{x}_1|)}\ , x_{2,i} = \frac{x_{2,i}}{\max(|\boldsymbol{x}_2|)} \text{ for } i = 1, \ldots, N \tag{3.39}$$

the data are restricted without loss of information to a domain $\{a_1 \mid 0 < a_1 \leq 1\}$, $\{a_2 \mid 0 < a_2 \leq 1\}$ known to the ML model.

Equation (2.11) shows that the coherence magnitude γ and the interferometric phase φ, which is optimally estimated by the sample estimator $\hat{\varphi}_\mathrm{s}$, are independent of each other. Hence, assuming a stationary phase signal, the expected interferometric phase φ can be estimated from the statistically homogeneous pixels (SHPs)

$$\begin{aligned} \varphi =& \arg\left(\mathrm{E}\{\boldsymbol{X}_1 \cdot \boldsymbol{X}_2^*\}\right) \\ \approx& \hat{\varphi}_\mathrm{s} = \arg\left(\sum_{i=1}^{N} x_{1,i} \cdot x_{2,i}^*\right). \end{aligned} \tag{3.40}$$

Since only the interferometric phase difference between each i.i.d. sample is used, the expected value can be compensated in the primary scene in advance

$$x_{1,i} = x_{1,i} \cdot \exp(-j\hat{\varphi}_\mathrm{s}). \tag{3.41}$$

This transformation eliminates the phase ambiguities $\phi_{k,i} + K\,2\pi$ and preserves the respective amplitude's Rayleigh PDF of the primary and secondary scene and the statistics of the interferometric phase differences (except for the mean).

As a result of the amplitude scaling and interferometric phase compensation, the number of possible input data has now been significantly reduced.

Data Representation: The encoder converts the CCG input data because the data representation has an impact on the performance of the model. This is due to the fact that there is practically no direct regression from the input variables to the output value. Inside the ML model, attributes that are not visible from the outside are calculated. This internal automatic generation of features is manually supported by the encoding. Two examples for possible CCG data representations are $\{\mathfrak{Re}(\boldsymbol{x}_1), \mathfrak{Im}(\boldsymbol{x}_1), \mathfrak{Re}(\boldsymbol{x}_2), \mathfrak{Im}(\boldsymbol{x}_2)\}$ and $\{|\boldsymbol{x}_1|, |\boldsymbol{x}_2|, \arg(\boldsymbol{x}_1\,\boldsymbol{x}_2^*)\}$. Tests have shown that the latter data representation, consisting of the sample amplitudes and expected interferometric phase compensated phase differences, is more advantageous than others.

ML Training

The ML training learns the features with which the internal model is evaluated to return the coherence magnitude estimate. According to James et al. (2021), the general form of ML regression is

$$\hat{\gamma} = f(.) + \epsilon. \tag{3.42}$$

ϵ is an inherent random error term and is named the irreducible error. In this application, it results from the random sampling and the limited sample size N but not from the noise in the data. Practically, every sample is differently representative and ϵ corresponds to unmeasured information which results in bias and variance of the coherence magnitude estimate. Touzi et al. (1999) have proven that an unbiased estimator, which is a function of the sample coherence magnitude, cannot be found. It follows that the ML estimator will also have a bias and a variance. In other words, ϵ is independent of the input data $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and can only be mitigated by increasing the sample size N reducing unmeasured information.

ML provides procedures for estimating $f(.)$ based on training data and approximately represents it by $\hat{f}(.)$. Depending on the ML method, $\hat{f}(.)$ is represented differently, such as a Decision Tree, a Random Forest or a Neural Network. According to James et al. (2021), the error from the approximation $\hat{f}(.)$ of a particular ML method is termed reducible error. It can be diminished by choosing an appropriate ML method and, if used, suitable Neural Network layers as well as optimizing the learning parameters such as the learning rate and the learning iteration count. In this thesis, Gradient Boosted Trees ML is implemented based on the XGBoost library with its C-API developed by Chen & Guestrin (2016), Chen et al. (2022).

All possible CCG input processes must be simulated for the ML training. To get as close to the real estimation scenario as possible, the amplitudes of the primary and secondary signals and the interferometric phases are modeled in such a way that the encoder works as it will later. In this thesis, the scenes' expected amplitudes are simulated with uniform likelihoods $a_1 \sim U(0, 2)$ and $a_2 \sim U(0, 2)$, and the expected interferometric phase with $\varphi \sim U(-\pi, \pi)$. For the training of an estimator, 10^8 interferograms are generated. In the course of the ML learning, the parameters of the model are tuned to perform best on the given training data. This suggests to add prior knowledge on the underlying coherence magnitude by adjusting the training data set. In doing so, the fact that ML learns the model from the data is exploited. Training data are generated with a number of observations corresponding to the prior on the underlying coherence. The assumption is that the ML parameter tuning then works better for these observed values than with the data, who has not or rarely seen the training. A single parameter γ_{max} is used to model the priors (3.5) and (3.6). In the following, this parameter is specified as a subscript at the respective method.

ML Without Prior (MLWP) Without prior information, training data are generated with the straight forward characteristic and γ_{simu} is sampled from the uniform distribution $\gamma_{\text{simu}} \sim U(0, 1)$.

ML Less Strict Prior (MLLSP) Figure 3.1b shows the distribution of γ_{simu} for the less strict prior. The implementation is based on the inverse CDF sampling method. It provides one random variate $\gamma_{\text{simu}} \sim \mathrm{p}_{\gamma}(\gamma)$ (3.6) from one random sample with distribution $u \sim U(0,1)$. The corresponding CDF is

$$\text{CDF} = \begin{cases} \frac{2\gamma}{\gamma_{max}+1} & 0 \leq \gamma \leq \gamma_{max} \\ \frac{(\gamma-2)\gamma+\gamma_{max}^2}{\gamma_{max}^2-1} & \gamma_{max} < \gamma \leq 1. \end{cases} \tag{3.43}$$

This leads to the respective inverse CDF

$$\text{CDF}^{-1} = \begin{cases} \frac{1}{2}(\gamma_{max}u + u) & u \leq \frac{2\gamma_{max}}{\gamma_{max}+1} \\ 1 - \sqrt{\gamma_{max}^2 u - \gamma_{max}^2 - u + 1} & u > \frac{2\gamma_{max}}{\gamma_{max}+1}. \end{cases} \tag{3.44}$$

ML Strict Prior (MLSP) Figure 3.1a visualizes the distribution of the underlying coherence magnitude γ for the strict prior. Consequently, the respective training data are generated with γ_{simu} sampled from the uniform distribution $\gamma_{\text{simu}} \sim U(0, \gamma_{max})$.

Practically, one ML model $\hat{f}_{N,p}(.)$ is generated for each prior type $\{p \mid \text{MLWP}, \text{MLLSP}_{\gamma_{\max}}, \text{MLSP}_{\gamma_{\max}}\}$ with every needed prior parameter $\{\gamma_{max} \mid 0.1, 0.2, \ldots, 0.9\}$. The utilized library XGBoost allows to persistently save each model into a JSON file (Chen & Guestrin 2016, Chen et al. 2022) for later operational estimation use.

3.2.2 Estimation of Coherence Magnitude

N interferometric samples are input to the operational coherence magnitude estimation. These data are transformed according to (3.39) and (3.41). I.e., $3 \times N$ real values, encoded by $\{|\boldsymbol{x}_1|/\max(|\boldsymbol{x}_1|), |\boldsymbol{x}_2|/\max(|\boldsymbol{x}_2|), \arg(\boldsymbol{x}_1\, \boldsymbol{x}_2^* \exp(-\mathrm{j}\hat{\varphi}_{\mathrm{s}})\}$, enter the ML prediction model $\hat{f}_{N,p}(.)$. Once again, all phases are the expected interferometric phase compensated phase differences. The model is extremely fast evaluated because no iteration, numeric integration or Bootstrapping is needed. The estimated coherence magnitude $\hat{\gamma}_p$ with $\{p \mid \text{MLWP}, \text{MLLSP}_{\gamma_{\max}}, \text{MLSP}_{\gamma_{\max}}\}$ is deterministic, i.e., one and the same input data result in one and the same estimate.

3.3 Composite Estimation

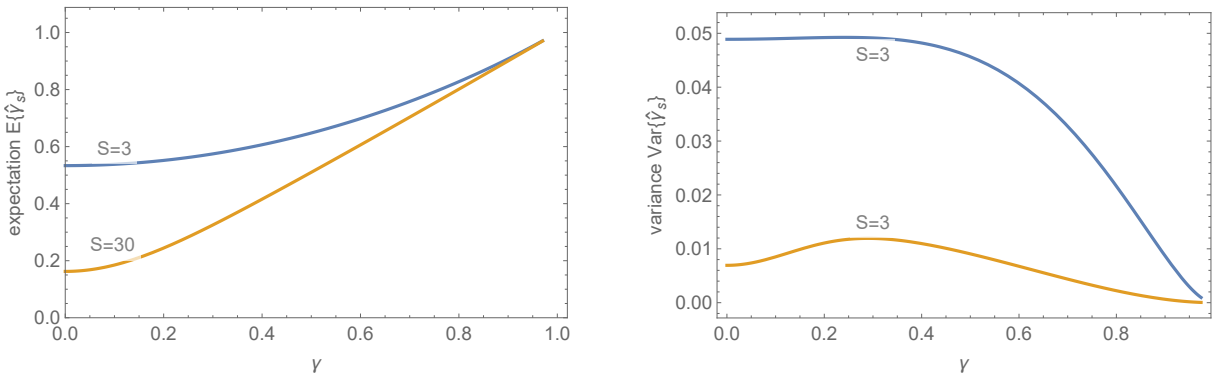
The method described in this section was originally published by the author in (Adam 2024).

Subject is to develop a new method based on the composition of coherence estimates from subsamples and ML. Starting point are N available i.i.d. InSAR sample pairs representing not perfectly the population of the random CCG processes of the primary $\boldsymbol{X}_1 : \{x_{1,1}, x_{1,2}, ..., x_{1,N}\}$ and secondary channel $\boldsymbol{X}_2 : \{x_{2,1}, x_{2,2}, ..., x_{2,N}\}$. As a result, all estimators of the coherence magnitude, e.g., the sample estimator (2.11), are biased, and the smaller the coherence and the number of available samples, the more biased they are. Figure 2.2 visualizes this effect by the conditional PDF, i.e., $\mathrm{p}_{\hat{\gamma}_\mathrm{s} \mid \gamma}(\hat{\gamma}_\mathrm{s} \mid \gamma, N)$ provided in (2.12), of the estimated sample coherence $\hat{\gamma}_\mathrm{s}$ published by Carter et al. (1973) as well Touzi & Lopes (1996). Obviously, $N = 30$ samples represent the statistic of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ better than $N = 3$ samples and the underlying coherence can be estimated more precisely.

For large samples with size N, $P_{N,S}$ independent subsamples each of size S can be generated by partitioning the original sample

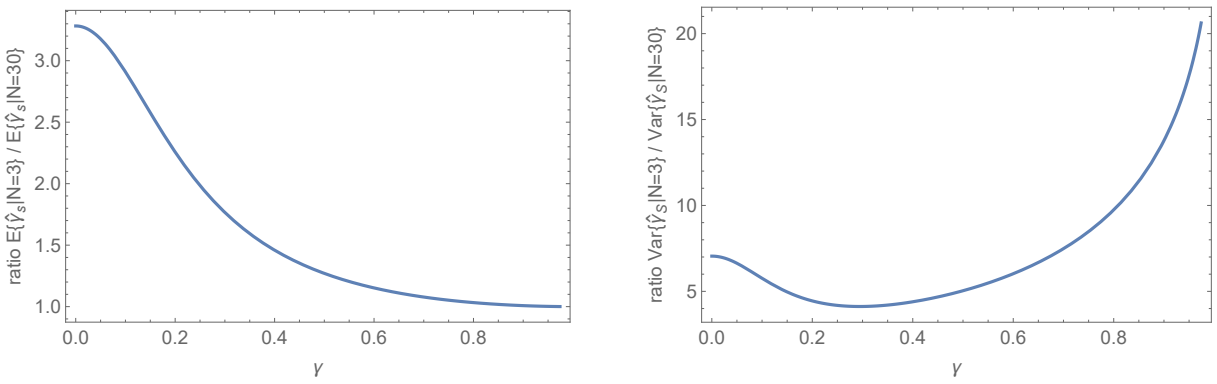
$$P_{N,S} = \lfloor \frac{N}{S} \rfloor. \tag{3.45}$$

$\lfloor . \rfloor$ is the *floor* function. For the example from Figure 2.2, the number of partitions is $P_{30,3} = 10$ and $P_{30,30} = 1$. The subsamples with respect to one and the same sample size S provide random, independent coherence estimates $\hat{\gamma}_1^{(S=3)}, ..., \hat{\gamma}_{10}^{(S=3)}$ (subsample group 1) and $\hat{\gamma}_1^{(S=30)}$ (subsample group 2). Each coherence estimate within a specific subsample group is distributed with a frequency of occurrence corresponding to the blue and orange PDF in Figure 2.2. Accordingly, the estimated values follow a specific statistic, whereby the statistic is different for unequal subsample sizes and estimation algorithms. Figure 3.8 visualizes examples of statistics related to the estimated coherence $\hat{\gamma}^{(S=3)}$ and $\hat{\gamma}^{(S=30)}$, i.e., from subsamples with $S = 3$ and $S = 30$ respectively for the expected value and variance of the sample estimator (2.11). Naturally, the estimates from the small subsamples



(a) Expectation $\mathrm{E}\{\hat{\gamma}_\mathrm{s} \mid N = S, \gamma = 0.3\}$. *(b)* Variance $\mathrm{Var}\{\hat{\gamma}_\mathrm{s} \mid N = S, \gamma = 0.3\}$.

Figure 3.8: Statistics of estimated coherence $\hat{\gamma}_\mathrm{s}$ given the true coherence $\gamma = 0.3$, $N = 3$ (blue) and $N = 30$ (orange) InSAR samples.



(a) Ratio of mean. *(b)* Ratio of variance.

Figure 3.9: Ratio of moments of the estimated coherence $\hat{\gamma}_\mathrm{s}$ from subsamples $S = 3$ and $S = 30$.

$\{\hat{\gamma}_i^{(S=3)} | i = 1, ..., P_{30,3}\}$ have a large variance as can be seen in Figure 3.8b by the blue graph. However, the variance can only be estimated using the numerous, in this example $P_{30,3} = 10$, independent subsample estimates. This means, the variance is introduced as a new feature useful for inference. For the single estimate $\hat{\gamma}^{(S=30)}$ the variance and higher moments describing the orange PDF in Fig. 2.2 are not available. Nevertheless, this estimate is the most precise we can obtain. As a result, many imprecise estimates provide valuable complementary information to the single complete sample estimate. Figure 3.9 shows examples of the ratio of statistical moments and it is obvious that the underlying coherence can be estimated from these ratios as features.

For the sample estimator, these relationships can still be derived for higher moments, however this is no longer possible for other estimators, as the estimation characteristics are not known as a formula in closed-form. In this thesis, ML is used for a practicable approach to derive the unknown functional relation directly from realistically simulated data. Essentially, the principle is to give the ML framework a set of independent coherence estimates with different estimation characteristics as features. It exploits the fact that ML can automatically evaluate the statistics of the subsample estimates for all possible estimators and thus makes inference for the coherence magnitude parameter estimation. In practice, it is a further development of the ML principle Adam (2023), as it is unable to evaluate the statistical relations described above. Similar to the previous ML method, this is a non-parametric approach and no knowledge of the statistical characteristics of the used estimators is required. This principle makes it possible to construct many different estimators, as numerous partitions with different subsample sizes can be generated and various estimators can be applied to them. Figure 3.10 visualizes the overall principle and workflow of the developed estimator with the configuration for $N = 30$ and subsample estimations $S = 3$ and $S = 30$ highlighted. In this figure, the partial estimators are abbreviated according to subsection 3.3.1 below. In practice, the algorithm is similar to a composite estimator which is an aggregation of different estimators (Lavancier & Rochet 2016). Instead of working with $N = 30$ complex interferometric samples $\{x_{1,i}, x_{2,i} | i = 1, ..., N\}$, independent coherence estimates, e.g., $P_{30,3} = 10$ times $\{\hat{\gamma}_i^{(S=3)} | i = 1, ..., 10\}$ and once $\hat{\gamma}_1^{(S=30)}$, are used for the training in the ML framework and for the estimation.

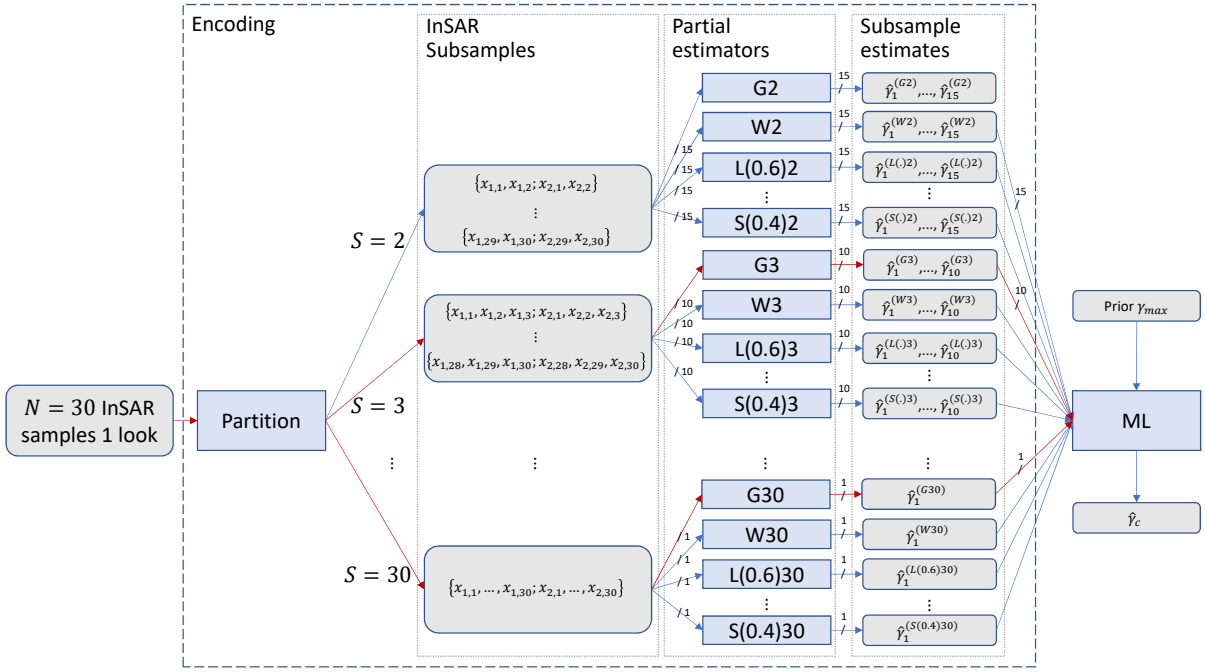


Figure 3.10: Composite estimator framework. The principle and workflow of the estimator *CW_N30_G3G30*, i.e., with the configuration $N = 30$ and subsample estimations $S = 3$ and $S = 30$, is highlighted.

3.3.1 Estimator Configuration Notation

For this thesis, a flexible framework was implemented according to Fig. 3.10. In order to refer to specific composite estimators that manifest themselves in a particular setup in this framework, the following notation is introduced. It is specified in Backus–Naur form (BNF).

```
<Setup> ::= 'C'<Prior>'_'<Looks>'_'<EstimatorList>
<Prior> ::= 'W' | <LSP> | <SP>
<LSP> ::= 'L'<MaxCoherence>
<SP> ::= 'S'<MaxCoherence>
<MaxCoherence> ::= '('<float>')'
<Looks> ::= 'N'<integer>
<EstimatorList> ::= <Estimator> |
                    <EstimatorList><Estimator>
<Estimator> ::= <SampleEstimator> | <MLEstimator>
<SampleEstimator> ::= 'G'< SubSampleSize>
<MLEstimator> ::= <MLWP> | <MLLSP> | <MLSP>
<MLWP> ::= 'W'<SubSampleSize>
<MLLSP> ::= <LSP><SubSampleSize>
<MLSP> ::= <SP><SubSampleSize>
<SubSampleSize> ::= <integer>
<integer> ::= <digit> | <integer><digit>
<float> ::= '0.'<integer>
<digit> ::= '0'|'1'|'2'|'3'|'4'|'5'|'6'|'7'|'8'|'9'
```

The first C in the name stands for composite estimator. This identifier is followed by an entry for the prior of the composite estimator. Accordingly, this composite estimator was trained with this prior and the input InSAR data correspond to this prior during the estimation. The number of InSAR sample pairs is coded next with the identifier N and is surrounded by underscores. This number is required to determine the number of partitions $P_{N,S}$ (3.45). In this thesis, compositions of the sample estimator abbreviated G, and machine learning estimators without prior named W or with less strict prior named $L(\gamma_{max})$ or strict prior named $S(\gamma_{max})$ are assessed for different configurations of subsamples and partitions. For example, CW_N30_G3G30 corresponds to the composite estimator (C) without prior (W) working with an InSAR sample size ($N30$) and aggregating all possible estimates with subsample size $S = 3$ ($G3$) and $S = 30$ ($G30$), i.e., $\{\hat{\gamma}_i^{(G3)} | i = 1, ..., 10\}$ and $\hat{\gamma}_1^{(G30)}$. Figure 3.11 visualizes the work- and dataflow of this specific configuration. Other composite estimators used in this thesis are CW_N30_G2, CW_N9_G2G9, $CL(0.6)\_N9\_L(0.6)2G9$, $CS(0.6)\_N30\_S(0.6)2G30$, $CS(0.6)\_N30\_S(0.6)2S(0.6)30$ and $CW\_N60\_G2G30G60$. CW_N30_G2 implements the inference without prior based on $P_{30,2} = 15$ independent estimates from the sample estimator ($G2$). CW_N9_G2G9 represents a composite estimation without prior from $P_{9,2} = 4$ sample estimates ($G2$) and a single ($P_{9,9} = 1$) sample estimate ($G9$). The estimator $CL(0.6)\_N9\_L(0.6)2G9$ is trained with a less strict prior ($L(0.6)$), i.e., with $\gamma_{max} = 0.6$ and does inference from $P_{9,2} = 4$ machine learning less strict prior estimates ($L(0.6)2$) and a single ($P_{9,9} = 1$) sample coherence estimate ($G9$).

A more complex example of this notation is $CW\_N30\_G3G30L(0.6)3$. It serves to explain an important principle regarding the input data of the partial estimators. In this particular configuration, the sample estimators $G3$ and the machine learning less strict prior esti-

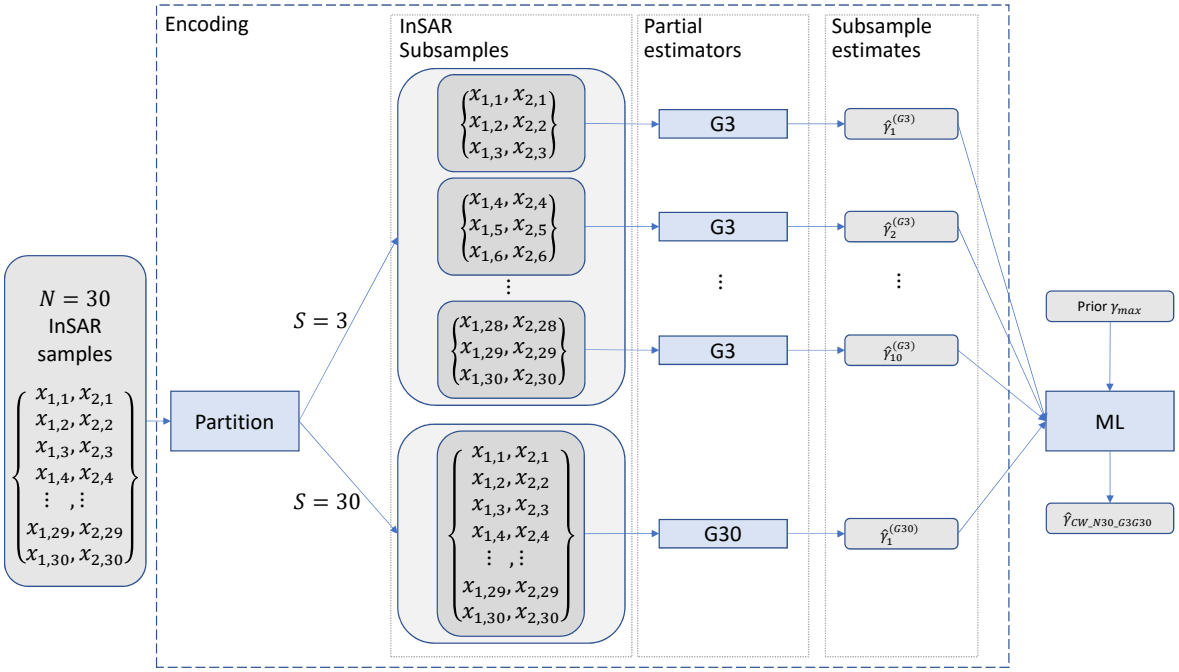


Figure 3.11: Detailed principle and workflow of the estimator CW_N30_G3G30.

mators $L(0.6)3$ are working on one and the same partition, i.e., all partial estimators have corresponding InSAR input samples. Figure 3.10 shows that the partition is fix for all partial estimators with a given subsample size.

According to the label for the partial sample estimator, the identifier G is in this thesis also used to refer to the corresponding standalone sample estimator. Hence, the abbreviation $G30$ corresponds to the sample estimator (2.11) with $N = 30$ InSAR samples. However, the abbreviations for the standalone estimators for machine learning without prior (MLWP) and for machine learning with strict prior ($\text{MLSP}_{\gamma_{max}}$) are retained corresponding to section 3.2.

3.3.2 Detailed Implementation

The implementation is more complex than the ML method from section 3.2, since these estimators are subcomponents, as Fig. 3.10 shows. In other words, the partial machine learning algorithms named W, $L(\gamma_{max})$ and $S(\gamma_{max})$ are practically implemented by MLWP, $\text{MLLSP}_{\gamma_{max}}$ and $\text{MLSP}_{\gamma_{max}}$ respectively. For the implementation, Fig. 3.12 visualizes the workflow and the functional decomposition into subcomponents. Data are shown in gray and subfunctions in blue boxes. Arrows represent the dataflow.

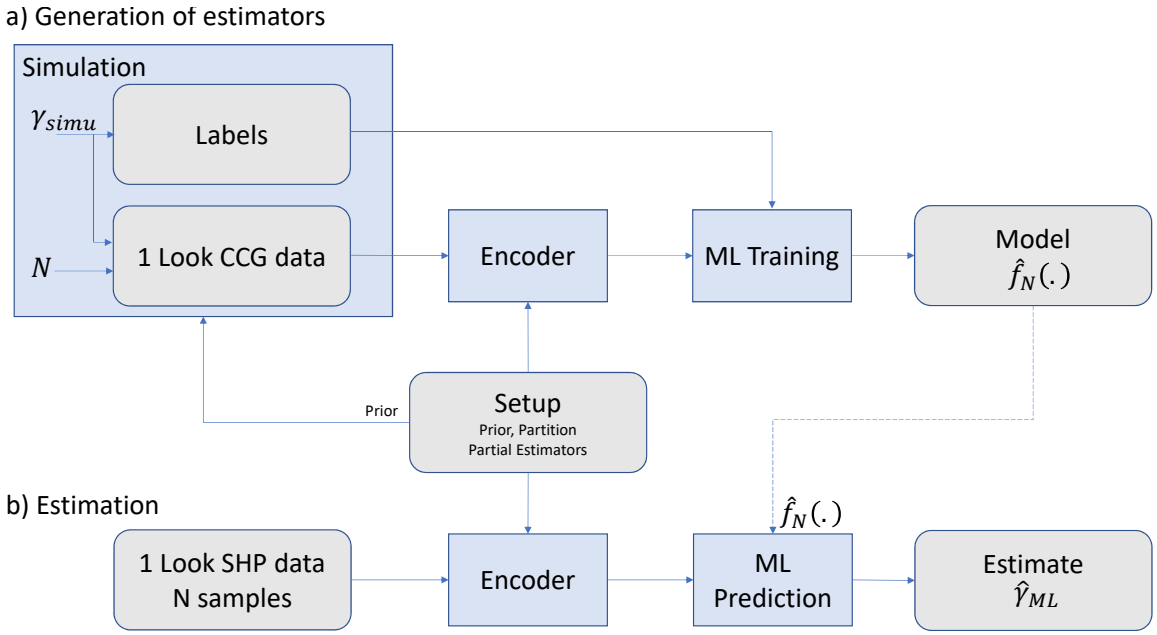


Figure 3.12: Principle of the ML framework for coherence magnitude estimation.

Simulation

The simulation of the data and the prior are based on the same principle as described in section 3.2.1. A covariance matrix $\mathbf{\Sigma}$ is setup according to (3.30) with simulation parameters expected amplitudes a_1 and a_2 which are the primary and secondary scene's amplitude, InSAR phase φ and coherence magnitude γ. The simulation is based on random values with uniform likelihoods $a_1 \sim U(0, a_{max})$ and $a_2 \sim U(0, a_{max})$, and the interferometric phase with $\varphi \sim U(-\pi, \pi)$. In case of γ, the simulation PDF corresponds to the prior of the composite estimator $\gamma \sim prior(\gamma_{max}) = \mathrm{p}_\gamma(\gamma)$ as described in respective paragraphs ML Without Prior (MLWP), ML Less Strict Prior (MLLSP) and ML Strict Prior (MLSP). Also, the actual γ from the random process is used as the training label. Next, the covariance matrix $\mathbf{\Sigma}$ is transformed into a random InSAR data sample $\boldsymbol{S}$ using (3.37).

Encoding

The encoding (Fig 3.10) consists of the partition algorithm and the partial estimators. In Fig. 3.13, the partition subfunction is detailed for a sample size $N = 9$. The demonstrated principle extends to other sample sizes accordingly. In particular, the partitioning is fixed and unchangeable. Subsamples are always created that are independent inside a subsample group. These subsamples are processed by coherence magnitude estimators which are to be configured. In this thesis the sample estimator (2.11), and the machine learning algorithms MLWP, $\mathrm{MLLSP}_{\gamma_{max}}$ and $\mathrm{MLSP}_{\gamma_{max}}$ from section 3.2 are implemented. The respective prior needs to be configured for the simulation in the course of the training as well.

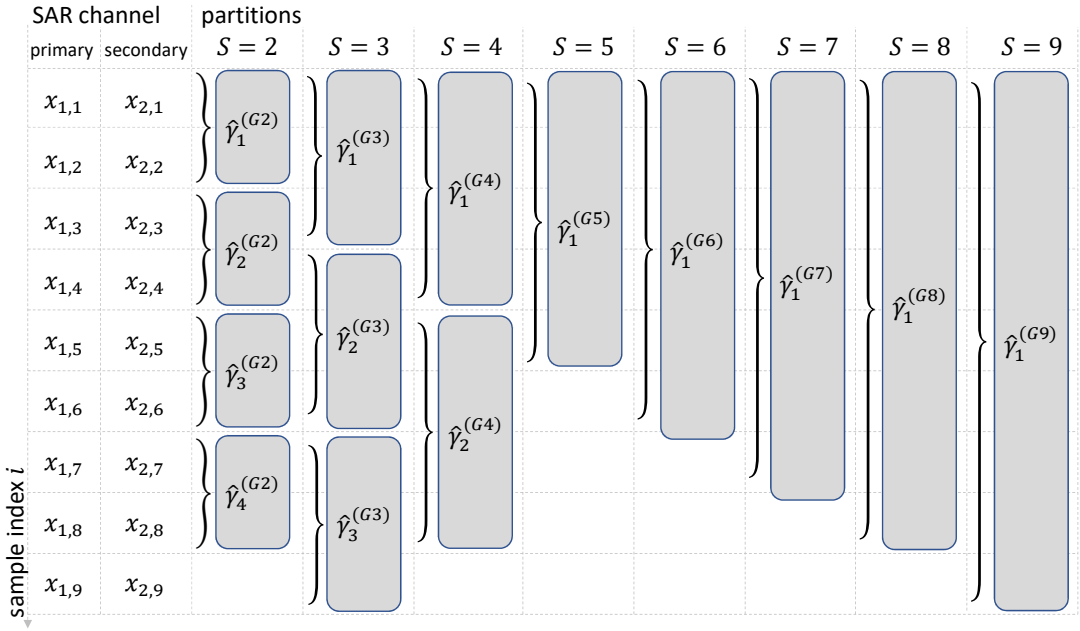


Figure 3.13: Principle of the partition subfunction for $N = 9$.

ML Training

The ML training is based on the same principle as described in section 3.2.1. Technically, it is not necessary for the partial estimators to be implemented with the same framework as the composite estimators. However, in this thesis, Gradient Boosted Trees ML is also implemented based on the XGBoost library with its C-API developed by Chen & Guestrin (2016) and Chen et al. (2022). For the training of an estimator, 10^8 InSAR samples $\boldsymbol{S}$ are generated using (3.37).

Estimation

In contrast to the direct estimation described in section 3.2, the input data are not rescaled and the expected interferometric phase is also not compensated. Instead, the sample is partitioned and the subsamples are processed by the respective configured estimators. Finally, ML evaluates the trained regression function $\hat{f}(.)$ with the partial estimates as arguments. Figure 3.10 visualizes this principle in the dashed box.

4 Experimental Results

The characteristics of the developed estimators are evaluated numerically by simulations without simplification or modeling. All estimators are applied on one and the same simulated samples from CCG processes described in section 3.2.1 in paragraph Simulation. Because of its practical importance, the conventional sample estimator $\hat{\gamma}_{\mathrm{s}}$ (2.11) is taken as the reference and included in the comparison of the estimators. In practice, the coherence bias is extreme when the sample size is small. This is the reason why these problematic scenarios are given priority in the assessment. In order to compare the developed estimators and demonstrate the complete characteristics, the bias

$$\gamma_*^{(\mathrm{bias})} = \mathrm{E}\{\hat{\gamma}_* - \gamma_{true}\} \tag{4.1}$$

and the standard deviation

$$\gamma_*^{(\sigma)} = \sqrt{\mathrm{E}\{(\hat{\gamma}_* - \mathrm{E}\{\hat{\gamma}_*\})^2\}} \tag{4.2}$$

as well as the RMSE

$$\gamma_*^{(\mathrm{RMSE})} = \sqrt{\mathrm{E}\{(\hat{\gamma}_* - \gamma_{true})^2\}} \tag{4.3}$$

are experimentally estimated by simulations because these are the relevant quality criteria of estimators. $\hat{\gamma}_*$ is the estimate from the respective estimator presented in section 3.1.3 for $\{\hat{\gamma}_* \mid \hat{\gamma}_{\mathrm{MAP}}, \hat{\gamma}_{\mathrm{EAP}}, \hat{\gamma}_{\mathrm{EAPLSP}_{\gamma\mathrm{max}}}, \hat{\gamma}_{\mathrm{EAPSP}_{\gamma\mathrm{max}}}, \hat{\gamma}_{\mathrm{MEDAP}}\}$, in section 3.2 for $\{\hat{\gamma}_* \mid \hat{\gamma}_{\mathrm{MLWP}}, \hat{\gamma}_{\mathrm{MLLSP}_{\gamma\mathrm{max}}}, \hat{\gamma}_{\mathrm{MLSP}_{\gamma\mathrm{max}}}\}$ and in section 3.3 for $\{\hat{\gamma}_* \mid \hat{\gamma}_{CW\_N\ldots}, \hat{\gamma}_{CL(\gamma\mathrm{max})\_N\ldots}, \hat{\gamma}_{CS(\gamma\mathrm{max})\_N\ldots}\}$. The general quality criterion is the RMSE because it combines the bias (4.1) and the standard deviation (4.2) into a single quality metric by the relation

$$\gamma_*^{(\mathrm{RMSE})} = \sqrt{\left(\gamma_*^{(\sigma)}\right)^2 + \left(\gamma_*^{(\mathrm{bias})}\right)^2}. \tag{4.4}$$

In this thesis, the RMSE is used instead of the MSE for the characterization of the estimator efficiency because this parameter is in the same scale as the true coherence.

As a proof of concept, the estimators are demonstrated also using real Sentinel-1 data in Interferometric Wide swath mode. One and the same test site is used for all estimators. The primary scene has the orbit number 30741 and was acquired on January 10, 2020. After 12 days, the secondary scene was recorded. Their orbit number is 30916 and the observation geometry is characterized by an effective baseline of about 27 m. Without going into details, the oversampling in the input data is reversed and the estimation window of 3×3 samples in range and azimuth (i.e., $N = 9$) overlaps from sample to sample.

4.1 Empirical Bayesian Estimation

4.1.1 Comparison of MAP, EAP and MEDAP Estimators

This experiment compares the three basic types of Bayesian estimators MAP (3.8), EAP (3.9) and MEDAP (3.10) which are practically implemented by (3.25), (3.26) and (3.27) respectively. The sample sizes $N = 3$ and $N = 9$ are chosen to check for the best method, which is analyzed in more detail below. Figure 4.1 visualizes the comparison results for $N = 3$ in the left and for $N = 9$ in the right column. The respective bias is provided in Fig. 4.1a and Fig. 4.1b and the standard deviation in Fig. 4.1c and Fig. 4.1d.

All newly developed estimators, are bias free at much smaller coherences. The EAP is bias free around a true coherence of 0.46, the MEDAP around 0.54 and the MAP around 0.8. However, after that, the bias for the EAP and MEDAP estimators becomes notably negative.

Figure 4.1e shows for $N = 3$ that EAP is up to 0.54, the MEDAP method up to 0.53 and the MAP estimator up to 0.47 more efficient than the conventional sample estimator for all underlying coherences.

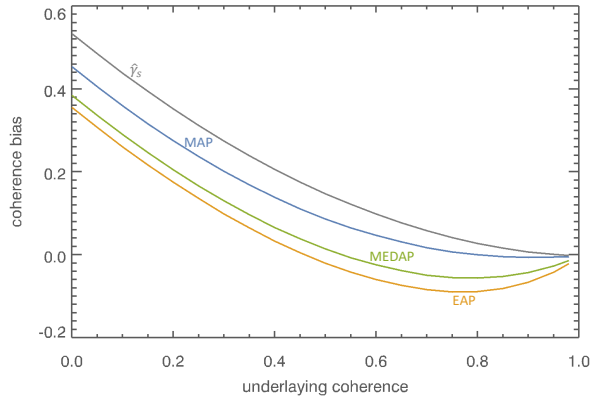
Figure 4.1f visualizes the overall performance characteristic for each estimator with $N = 9$. Here, the EAP and the MEDAP methods are more efficient compared to the conventional sample estimator for all underlying coherences up to 0.37 and 0.36 accordingly. And the MAP is more efficient for true coherences up to 0.31.

4.1.2 Comparison of EAP, EAPLSP and EAPSP Estimators

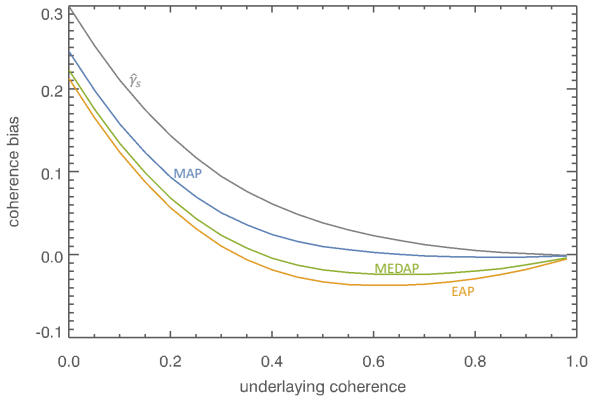
The experiment in section 4.1.1 demonstrates that the EAP estimator performs best compared to the MAP and MEDAP estimators. For this reason, the characteristics of the of the EAP estimator are further evaluated and compared with the variants with an informative prior, i.e., the EAPLSP and EAPSP estimators. Figure 4.2 compares the characteristics for $N = 3$ and $N = 9$ on the left and the right column respectively. As an example, the maximum underlying coherence is set to $\gamma_{\max} = 0.6$. The respective general priors are plotted in orange in Fig. 3.6.

The bias plots in Figs. 4.2a and 4.2b show that for small coherences the bias is reduced compared to the sample estimator and becomes smaller the stricter the prior is. The effect is all the more pronounced the smaller N is. Therefore, some numbers for $N = 3$ follow. For a zero coherence, the bias is reduced from 0.534 to 0.26 (51.3 %) for the $\text{EAPSP}_{0.6}$, to 0.338 (36.6 %) for the $\text{EAPLSP}_{0.6}$ and to 0.356 (33.3 %) for the EAP estimator with uninformative prior.

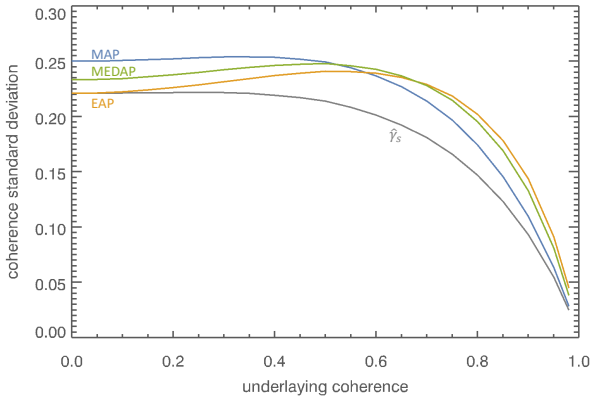
For the sample estimator, the bias becomes zero at a true coherence of one. Once again, all newly developed estimators, are bias free at much smaller coherences as well compared to the uninformative setup. For example, for $N = 3$, the $\text{EAPSP}_{0.6}$ is bias free around a true coherence of 0.27, the $\text{EAPLSP}_{0.6}$ around 0.42 and the EAP around 0.46.



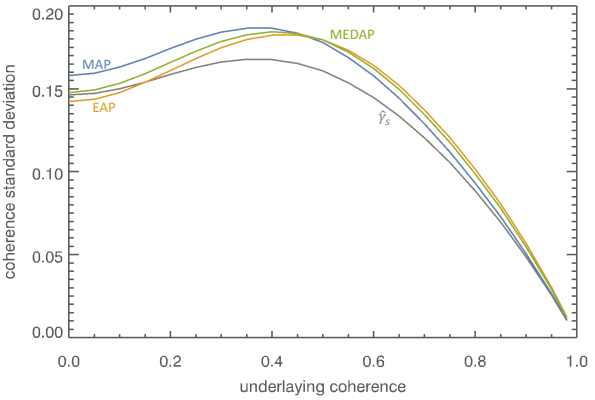
(a) Bias $\gamma_*^{(\mathrm{bias})}$ for $N = 3$.



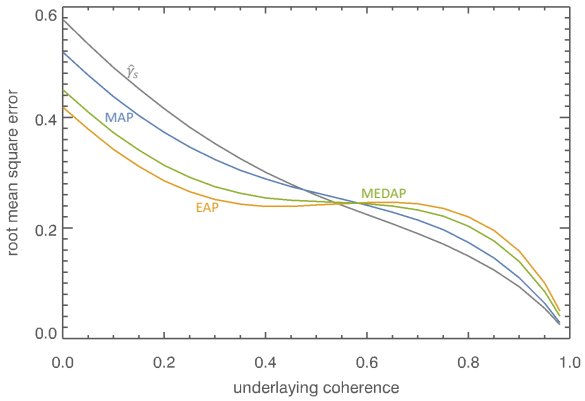
(b) Bias $\gamma_*^{(\mathrm{bias})}$ for $N = 9$.



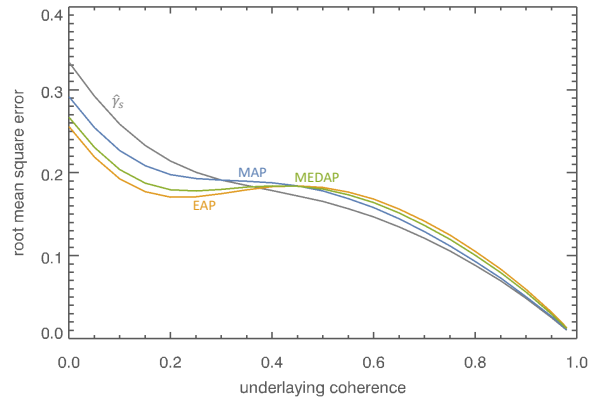
(c) Standard deviation $\gamma_*^{(\sigma)}$ for $N = 3$.



(d) Standard deviation $\gamma_*^{(\sigma)}$ for $N = 9$.

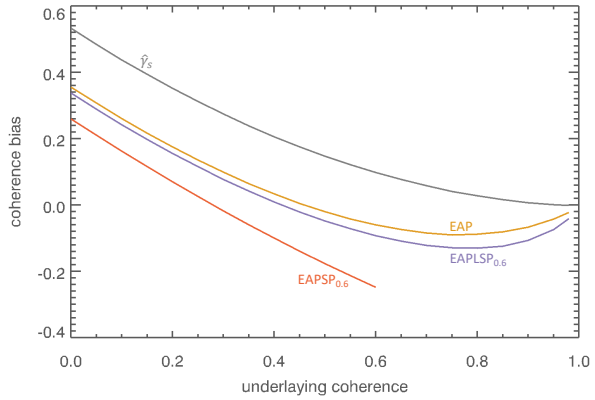


(e) RMSE $\gamma_*^{(\mathrm{RMSE})}$ for $N = 3$.

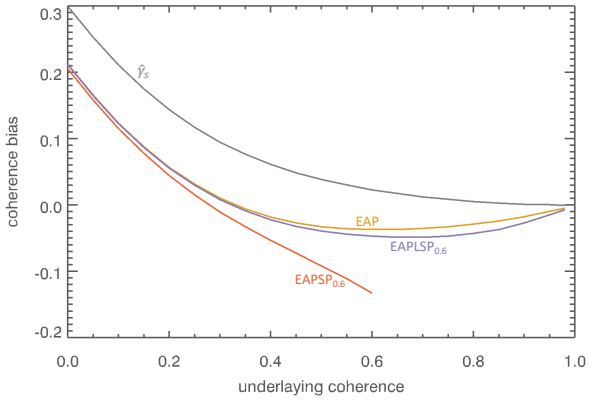


(f) RMSE $\gamma_*^{(\mathrm{RMSE})}$ for $N = 9$.

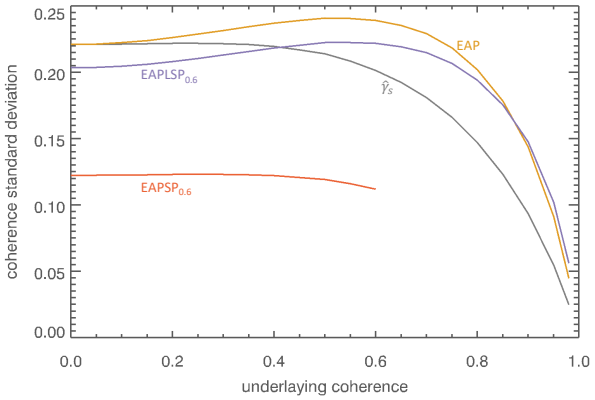
Figure 4.1: Characteristic of basic empirical Bayesian estimators MAP (blue), EAP (orange) and MEDAP (green) all without prior for $N = 3$ (left column) and $N = 9$ (right column) compared with the sample estimator $\hat{\gamma}_\mathrm{s}$ (gray).



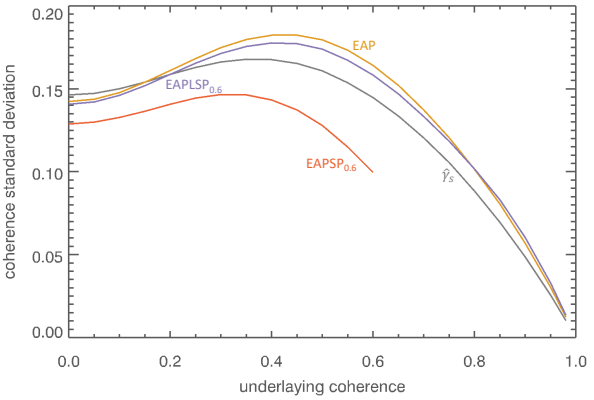
(a) Bias $\gamma_*^{(\text{bias})}$ for $N = 3$.



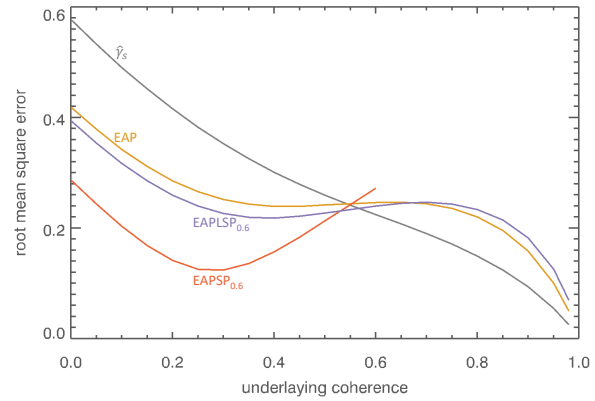
(b) Bias $\gamma_*^{(\text{bias})}$ for $N = 9$.



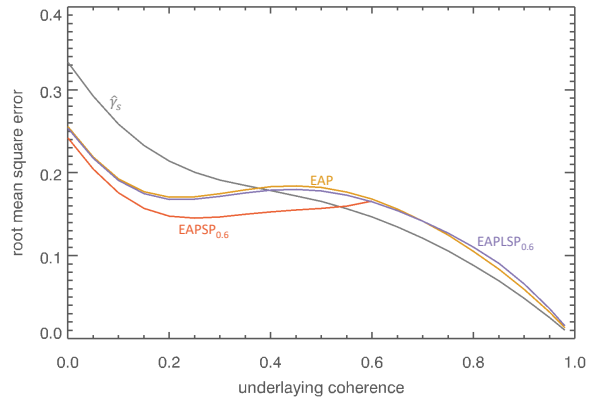
(c) Standard deviation $\gamma_*^{(\sigma)}$ for $N = 3$.



(d) Standard deviation $\gamma_*^{(\sigma)}$ for $N = 9$.



(e) RMSE $\gamma_*^{(\text{RMSE})}$ for $N = 3$.



(f) RMSE $\gamma_*^{(\text{RMSE})}$ for $N = 9$.

Figure 4.2: Characteristics of EAP estimators without (orange) and with prior information, i.e., EAPLSP$_{0.6}$ (purple) and EAPSP$_{0.6}$ (red) for $N = 3$ (left column) and $N = 9$ (right column) compared with the sample estimator $\hat{\gamma}_\text{s}$ (gray).

4.1.3 Characteristics for $N = 2$

It is well-known, the concept of coherence is not relevant to individual samples which implies $N \geq 2$. It results from the fact that the principle of ergodicity is used in (2.10) and the spatial average is used instead of the ensemble mean. Touzi & Lopes (1996) report that equation (2.12) is only valid for $N > 2$ samples. However, Carter et al. (1973) do not specify such a restriction for N. Therefore, in this section it is experimentally tested whether the Bayesian methods also work for $N = 2$ samples. Figure 4.3 presents for $N = 2$ the bias, standard deviation and RMSE for the estimators with uninformative (left column) and informative (right column) general priors. The simulation shows that all developed methods also work very well in this setup. Comparing the RMSE, especially the informative general priors are recommended for the estimation with only two independent InSAR samples.

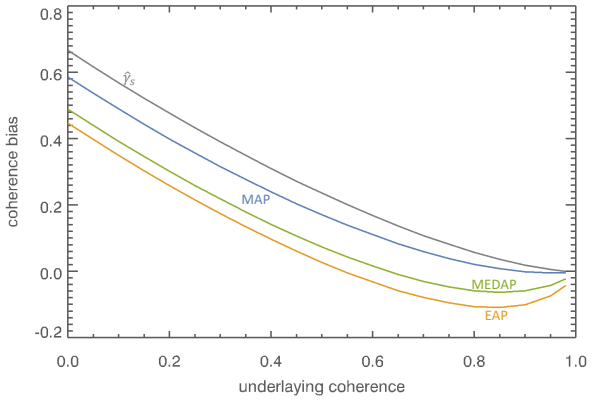
4.1.4 Characteristics for Large N

In the previous sections, the advantage of the Bayesian methods is demonstrated for small sample sizes. The characterization for large sample sizes is provided for $N = 15$ in Fig. 4.4, $N = 30$ in Fig. 4.5 and $N = 60$ in Fig. 4.6. The EAP estimators without prior are visualized in the left column and with informative prior in the right column. These examples show that even for large N, the bias of all methods is always smaller than the sample estimator's and still becomes negative. However, as the sample size increases, the bias vanishes and the estimators are also asymptotically unbiased. All EAP estimators except $\text{EAPSP}_{0.6}$ approach to a similar bias, standard deviation and RMSE characteristic. Especially, the RMSE indicates that the Bayes methods can be recommended for all sample sizes $N \leq 15$. It is noteworthy that the strict prior performs best also with large sample sizes. A limitation of the newly developed methods is that the performance improves only for low coherences if the sample size is small.

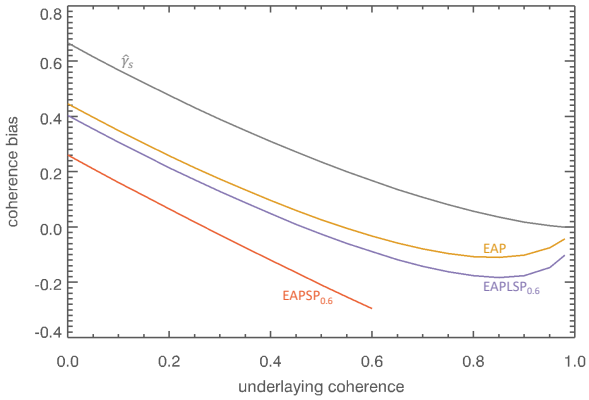
4.1.5 Sentinel-1 Application Demonstration

Figure 4.7a visualizes the test case with 512×512 i.i.d. samples by the radar backscatter amplitude. The coherence magnitude from the sample estimator is visualized in Fig. 4.7b. Using identical estimation windows, the respective EAP result is shown in Fig. 4.7d. In this example, the EAP coherence magnitude is estimated locally adaptive with respect to the prior from expected *a posteriori* without prior (EAPWP), $\text{EAPLSP}_{0.6}$, $\text{EAPLSP}_{0.4}$ or $\text{EAPSP}_{0.4}$. It can be seen that the estimation performance now depends not only on the window size but mainly on the prior and its strictness.

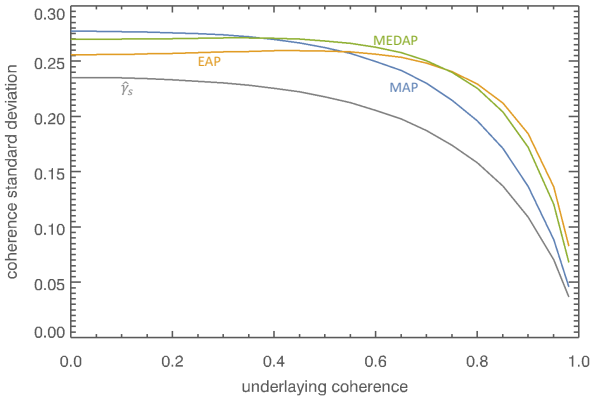
To give an intuitive idea of the effect of different priors and various parameters, Fig. 4.7c visualizes a composition of coherence estimates. In this figure from left to right, the result from the sample estimator, EAPWP, $\text{EAPLSP}_{0.6}$, $\text{EAPLSP}_{0.4}$, $\text{EAPSP}_{0.6}$, and $\text{EAPSP}_{0.4}$ can be compared. Similar coherence magnitudes are observed for all but the last two



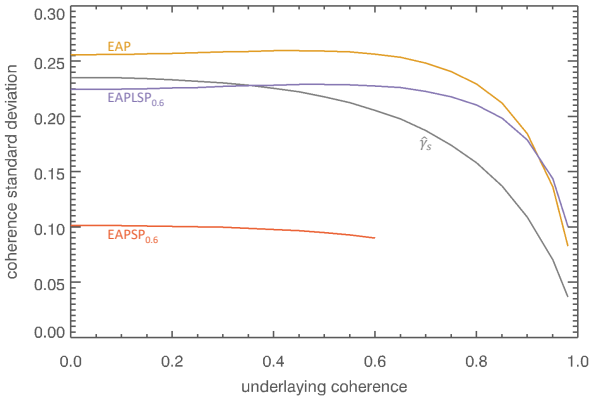
(a) Bias $\gamma_L^{(\text{bias})}$ without prior.



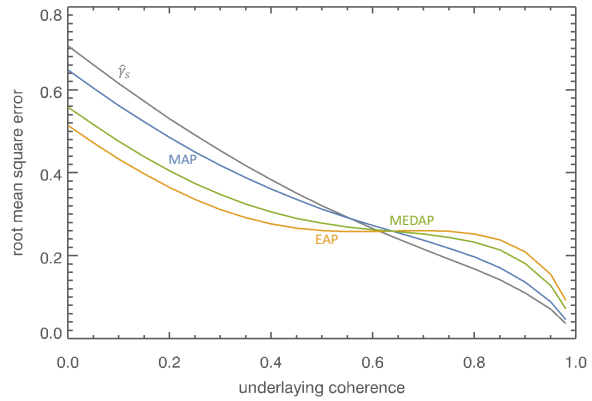
(b) Bias $\gamma_R^{(\text{bias})}$ informative priors.



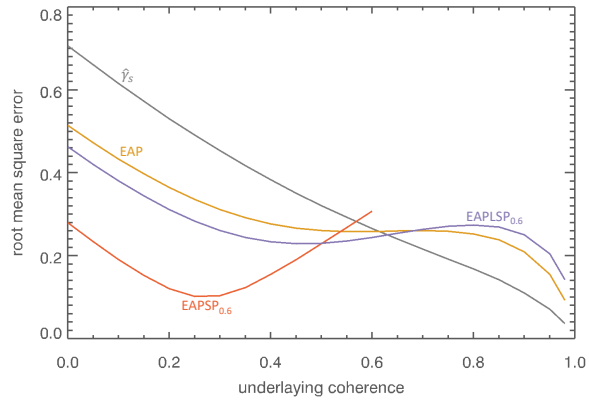
(c) Standard deviation $\gamma_L^{(\sigma)}$ without prior.



(d) Standard deviation $\gamma_R^{(\sigma)}$ informative priors.

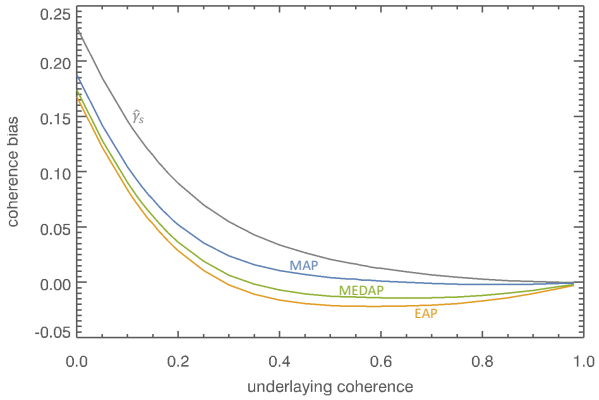


(e) RMSE $\gamma_L^{(\text{RMSE})}$ without prior.

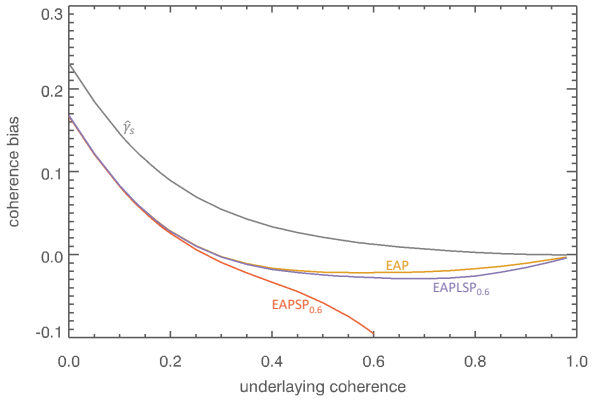


(f) RMSE $\gamma_R^{(\text{RMSE})}$ informative priors.

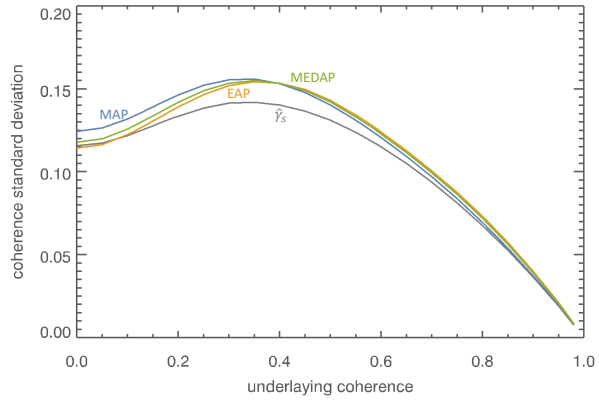
Figure 4.3: Characteristic of empirical Bayesian estimators for $N = 2$; left column provides the uninformative prior comparing $L \in \{\hat{\gamma}_\text{s}, \text{MAP}, \text{EAP}, \text{MEDAP}\}$ and the right column the informative prior comparing $R \in \{\text{EAPLSP}_{0.6}, \text{EAPSP}_{0.6}\}$ with $\hat{\gamma}_\text{s}$ and $\hat{\gamma}_\text{EAP}$.



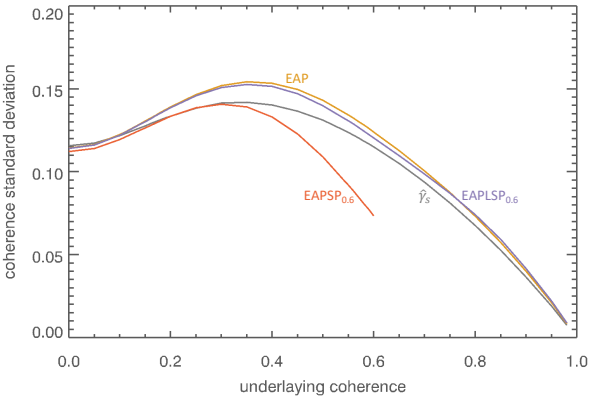
(a) Bias $\gamma_L^{(\text{bias})}$ without prior.



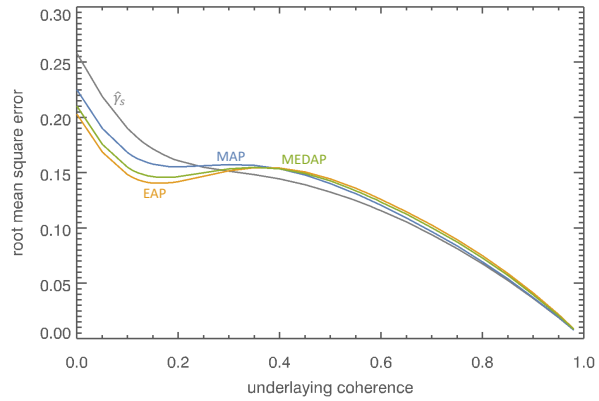
(b) Bias $\gamma_R^{(\text{bias})}$ informative priors.



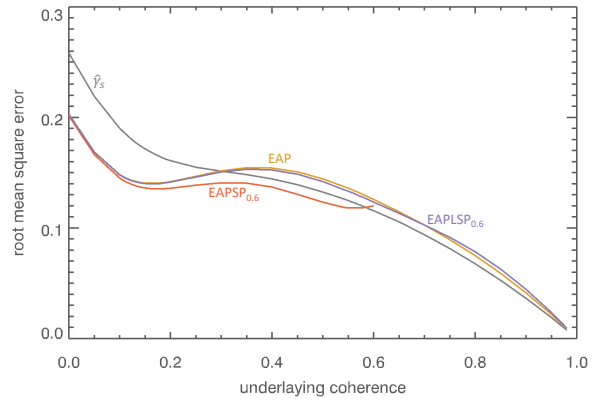
(c) Standard deviation $\gamma_L^{(\sigma)}$ without prior.



(d) Standard deviation $\gamma_R^{(\sigma)}$ informative priors.

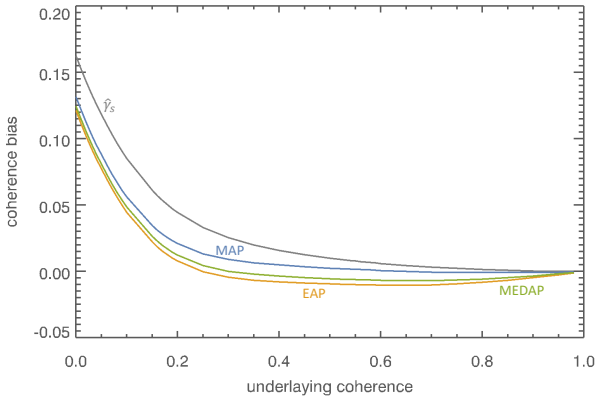


(e) RMSE $\gamma_L^{(\text{RMSE})}$ without prior.

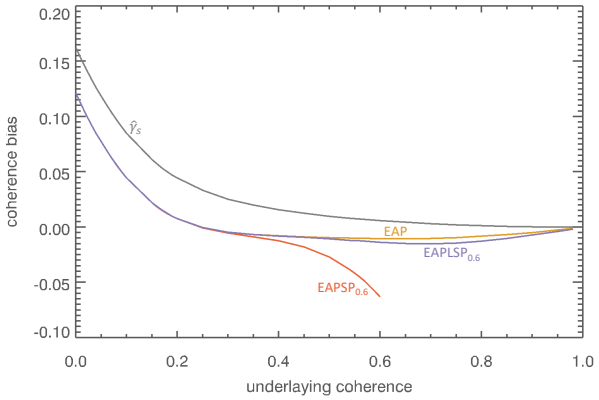


(f) RMSE $\gamma_R^{(\text{RMSE})}$ informative priors.

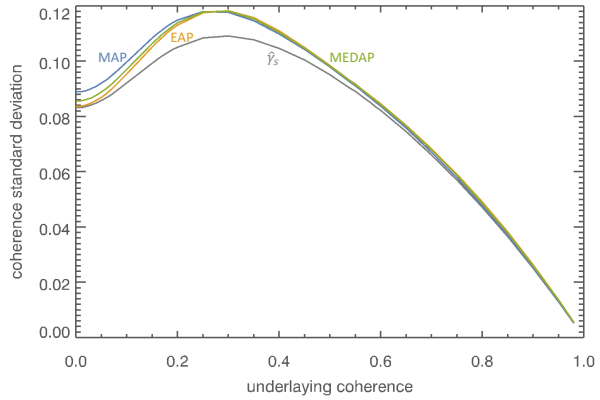
Figure 4.4: Characteristic of empirical Bayesian estimators for $N = 15$; left column provides the uninformative prior comparing $L \in \{\hat{\gamma}_\text{s}, \text{MAP}, \text{EAP}, \text{MEDAP}\}$ and the right column the informative prior comparing $R \in \{\text{EAPLSP}_{0.6}, \text{EAPSP}_{0.6}\}$ with $\hat{\gamma}_\text{s}$ and $\hat{\gamma}_\text{EAP}$.



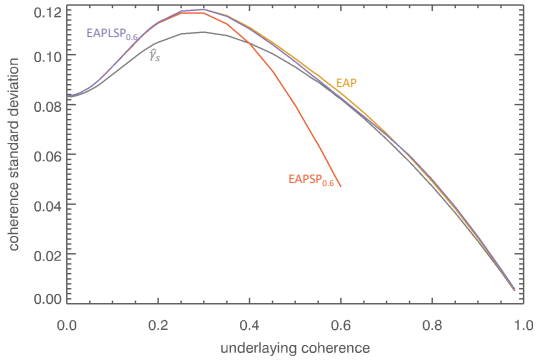
(a) Bias $\gamma_L^{(\text{bias})}$ without prior.



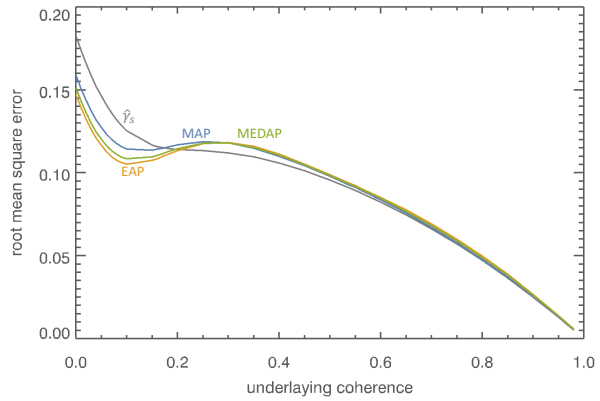
(b) Bias $\gamma_R^{(\text{bias})}$ informative priors.



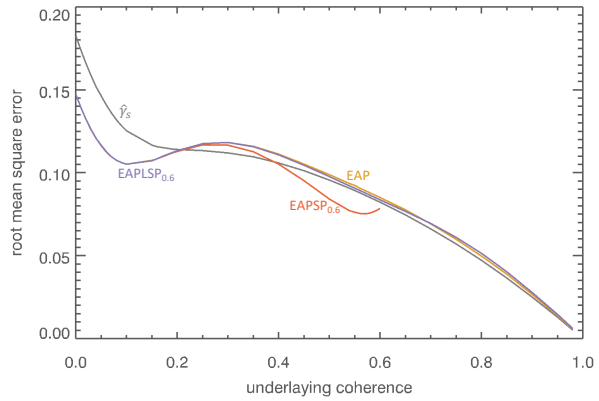
(c) Standard deviation $\gamma_L^{(\sigma)}$ without prior.



(d) Standard deviation $\gamma_R^{(\sigma)}$ informative priors.

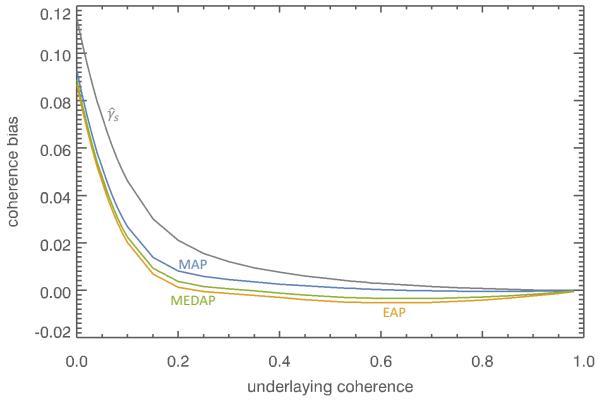


(e) RMSE $\gamma_L^{(\text{RMSE})}$ without prior.

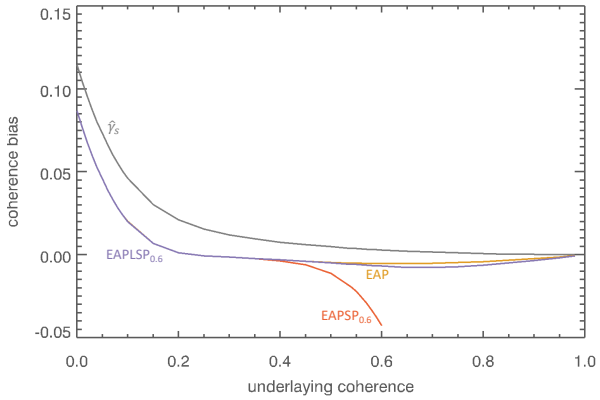


(f) RMSE $\gamma_R^{(\text{RMSE})}$ informative priors.

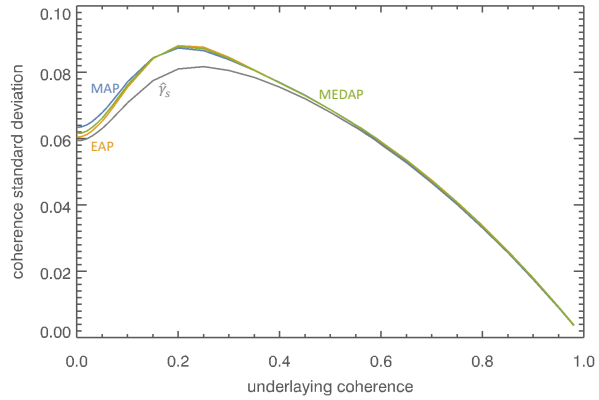
Figure 4.5: Characteristic of empirical Bayesian estimators for $N = 30$; left column provides the uninformative prior comparing $L \in \{\hat{\gamma}_\text{s}, \text{MAP}, \text{EAP}, \text{MEDAP}\}$ and the right column the informative prior comparing $R \in \{\text{EAPLSP}_{0.6}, \text{EAPSP}_{0.6}\}$ with $\hat{\gamma}_\text{s}$ and $\hat{\gamma}_\text{EAP}$.



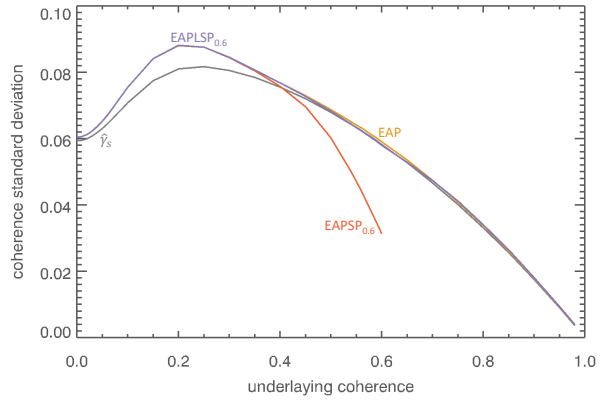
(a) Bias $\gamma_L^{(\text{bias})}$ without prior.



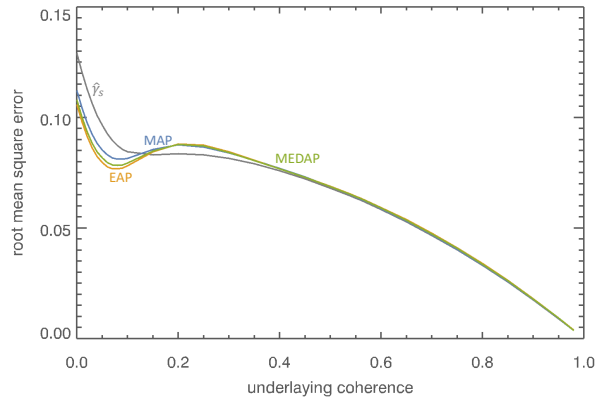
(b) Bias $\gamma_R^{(\text{bias})}$ informative priors.



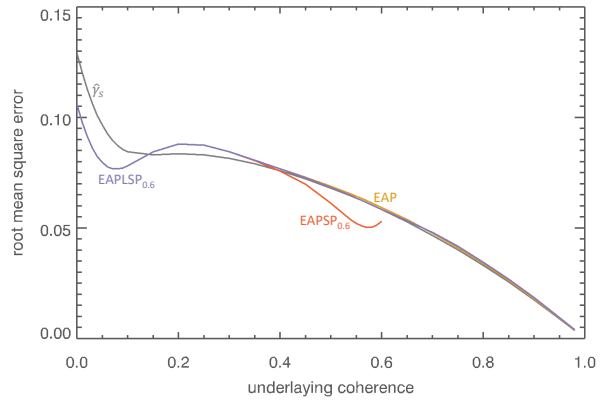
(c) Standard deviation $\gamma_L^{(\sigma)}$ without prior.



(d) Standard deviation $\gamma_R^{(\sigma)}$ informative priors.



(e) RMSE $\gamma_L^{(\text{RMSE})}$ without prior.



(f) RMSE $\gamma_R^{(\text{RMSE})}$ informative priors.

Figure 4.6: Characteristic of empirical Bayesian estimators for $N = 60$; left column provides the uninformative prior comparing $L \in \{\hat{\gamma}_\text{s}, \text{MAP}, \text{EAP}, \text{MEDAP}\}$ and the right column the informative prior comparing $R \in \{\text{EAPLSP}_{0.6}, \text{EAPSP}_{0.6}\}$ with $\hat{\gamma}_\text{s}$ and $\hat{\gamma}_\text{EAP}$.

columns. It follows that the less strict prior can robustly cope with an underlying coherence greater than the prior parameter $\gamma_{\max}$.

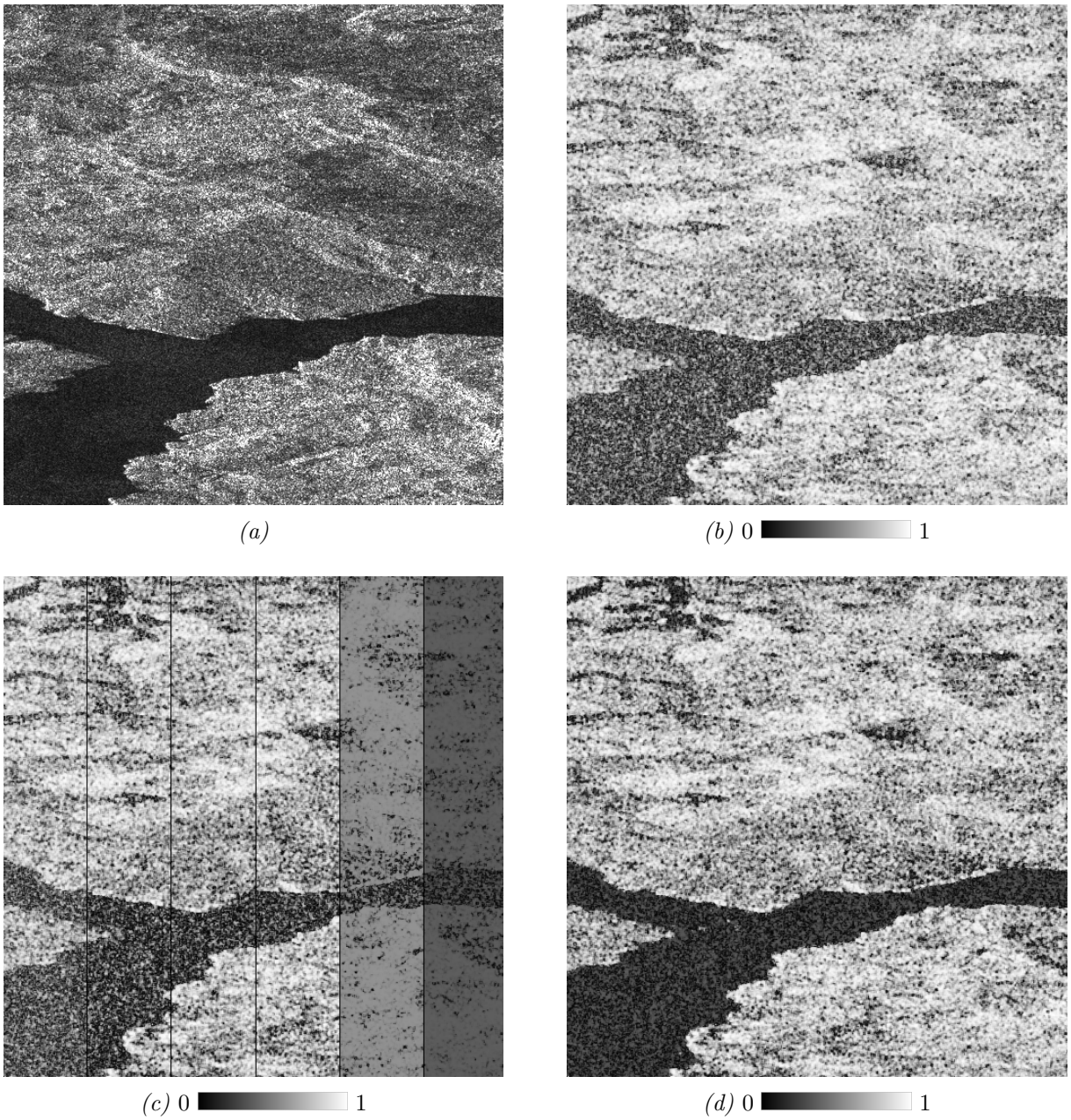


Figure 4.7: Sentinel-1 test case with $N = 9$ for Bayesian EAP estimators. (a) Radar backscatter amplitude. (b) Coherence magnitude from sample estimator $\hat{\gamma}_{\mathrm{s}}$. (c) EAP coherence estimates from sample estimator, EAP, $\mathrm{EAPLSP}_{0.6}$, $\mathrm{EAPLSP}_{0.4}$, $\mathrm{EAPSP}_{0.6}$, $\mathrm{EAPSP}_{0.4}$ from left to right. (d) EAP coherence magnitude estimated locally adaptive from EAPWP, $\mathrm{EAPLSP}_{0.6}$, $\mathrm{EAPLSP}_{0.4}$ or $\mathrm{EAPSP}_{0.4}$.

4.2 Machine Learning Estimation

4.2.1 Comparison of MLWP, MLLSP and MLSP Estimators

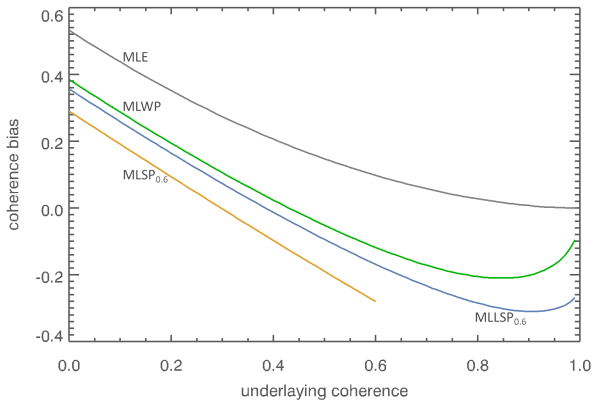
Based on the fact that the estimation from a small sample size is the critical problem, priority is put on such test cases, i.e., $N = 3$, and $N = 9$. The results in Fig. 4.8 show, the intuitively introduced Bayesian principle works. Any likelihood of prior can be implemented. In contrast to the empirical Bayesian approach described in section 3.1, no insoluble integral has to be solved and replaced by computationally ineffective numerical integration. Similar to the experiment in section 4.1.2, the prior parameter $\gamma_{\max} = 0.6$ is chosen because it is a typical value in InSAR.

In the plots below, the MLSP curves end at an underlying coherence of 0.6. It is apparent, a strict prior assumes zero probability outside of this range. However, it should be noted that the $\text{MLSP}_{0.6}$ estimator provides also estimates outside of this strict range.

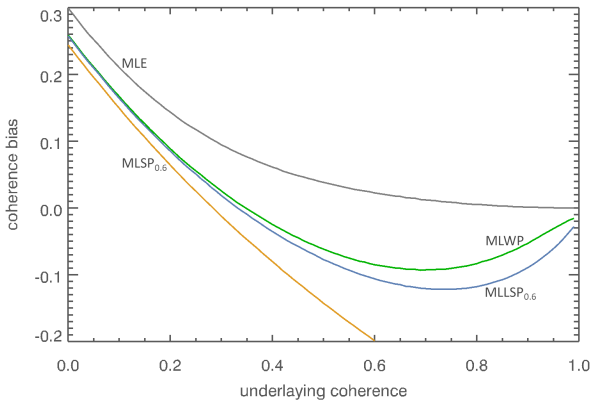
The bias compared in Figs. 4.8a and 4.8b is reduced for small coherences by all ML methods. For the sample estimator, the bias becomes zero at an underlying coherence of one. Not surprisingly, all newly developed ML estimators, are bias free at much smaller coherences. However, this is achieved at the expense of a larger bias for higher underlying coherence magnitude values.

The comparison of the RMSE in Figs. 4.8e and 4.8f confirms that the more information is used and the stricter the general prior, the more accurate the estimate will be. Compared to the conventional sample estimator, MLWP is more efficient for all underlying coherence magnitudes up to 0.68, the $\text{MLLSP}_{0.6}$ method up to 0.65, and the $\text{MLSP}_{0.6}$ estimator up to 0.58.

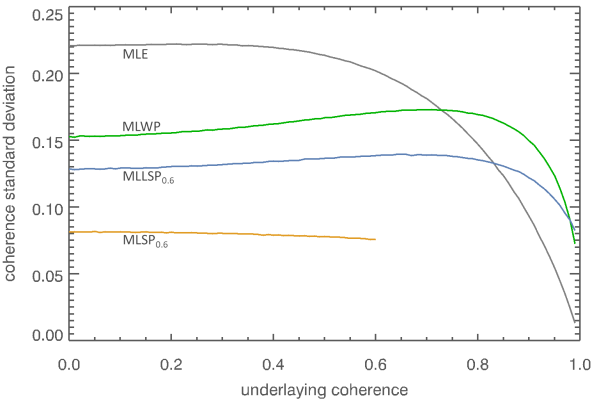
As the performance of the sample estimator improves with the number of samples, it can be expected that advantages are reduced for other methods. The comparison of the bias in Figs. 4.8a and 4.8b, of the standard deviation in Figs. 4.8c and 4.8d as well as of the RMSE in Figs. 4.8e and 4.8f confirm this expectation. Accordingly, the reduction in bias is less pronounced for $N = 9$. Also, the prior has less effect on the bias mitigation compared to test cases with fewer samples. A similar characteristic is observed for the standard deviation. However, some prior helps to mitigate the random variation. Nevertheless, the ML algorithms outperform the sample estimator for small coherence magnitude values in terms of RMSE. For $N = 3$ and compared to the conventional sample estimator, MLWP is more efficient for all underlying coherence magnitudes up to 0.62, the $\text{MLLSP}_{0.6}$ method up to 0.61 and the $\text{MLSP}_{0.6}$ estimator up to 0.55. For $N = 9$, the MLWP is more efficient for all underlying coherence magnitudes up to 0.43, the $\text{MLLSP}_{0.6}$ method up to 0.47 and the $\text{MLSP}_{0.6}$ estimator up to 0.48.



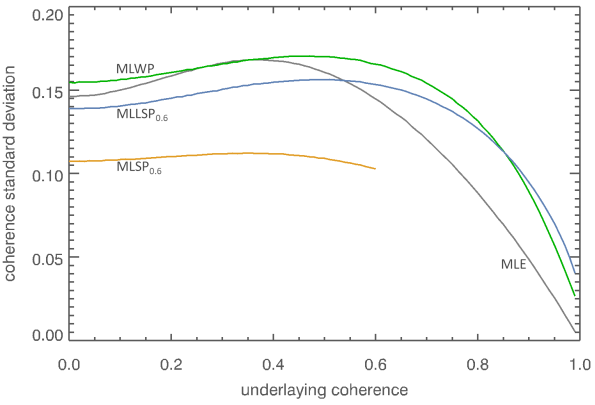
(a) Bias $\gamma_T^{(\mathrm{bias})}$ for $N = 3$.



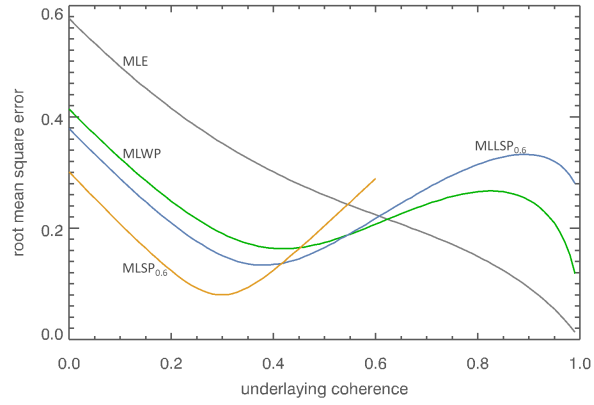
(b) Bias $\gamma_T^{(\mathrm{bias})}$ for $N = 9$.



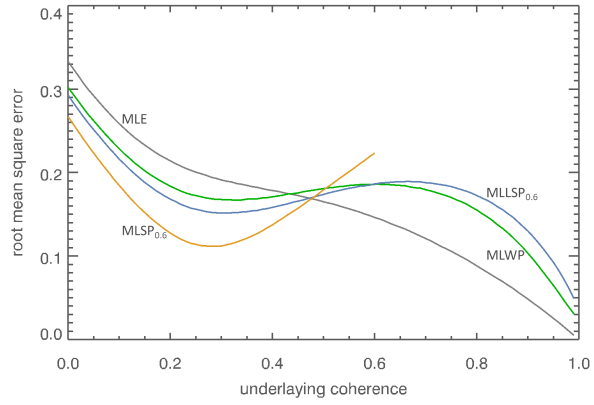
(c) Standard deviation $\gamma_T^{(\sigma)}$ for $N = 3$.



(d) Standard deviation $\gamma_T^{(\sigma)}$ for $N = 9$.



(e) RMSE $\gamma_T^{(\mathrm{RMSE})}$ for $N = 3$.

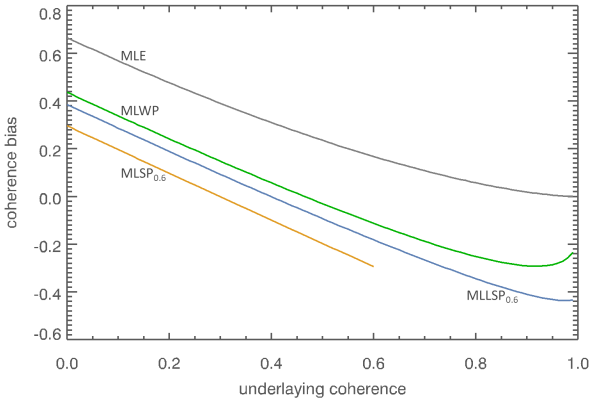


(f) RMSE $\gamma_T^{(\mathrm{RMSE})}$ for $N = 9$.

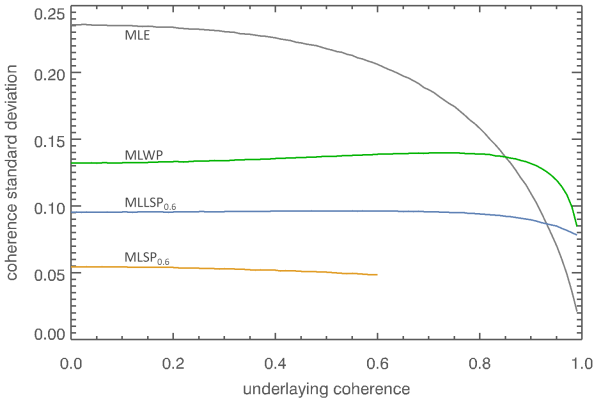
Figure 4.8: Characteristic of ML estimators for $N = 3$ left column and $N = 9$ right column comparing $T \in \{\hat{\gamma}_{\mathrm{s}}, \mathrm{MLWP}, \mathrm{MLLSP}_{0.6}, \mathrm{MLSP}_{0.6}\}$.

4.2.2 Characteristics for $N = 2$

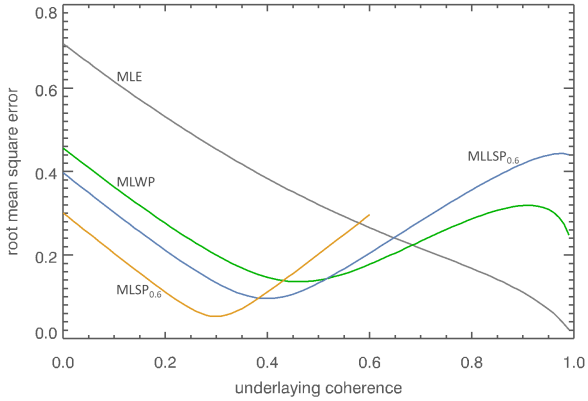
Similar to section 4.1.3, it is experimentally tested whether the ML methods also work for $N = 2$ samples. The results for this test case are visualized for the bias in Fig. 4.9a, the standard deviation in Fig. 4.9b and the RMSE in Fig. 4.9c. The simulation shows that all developed methods also work very well in this setup. The RMSE graphs in Fig. 4.9c confirm that the more information is used and the stricter the general prior, the more accurate the estimate will be. Compared to the conventional sample estimator, MLWP is more efficient for all underlying coherence magnitudes up to 0.68, the $\text{MLLSP}_{0.6}$ method up to 0.65, and the $\text{MLSP}_{0.6}$ estimator up to 0.58.



(a) Estimation bias $\gamma_*^{(\text{bias})}$.



(b) Standard deviation $\gamma_*^{(\sigma)}$.



(c) RMSE $\gamma_*^{(\text{RMSE})}$.

Figure 4.9: Characteristic of estimators for $N = 2$ samples; gray: sample estimator, green: MLWP, blue: $\text{MLLSP}_{0.6}$ and orange: $\text{MLSP}_{0.6}$.

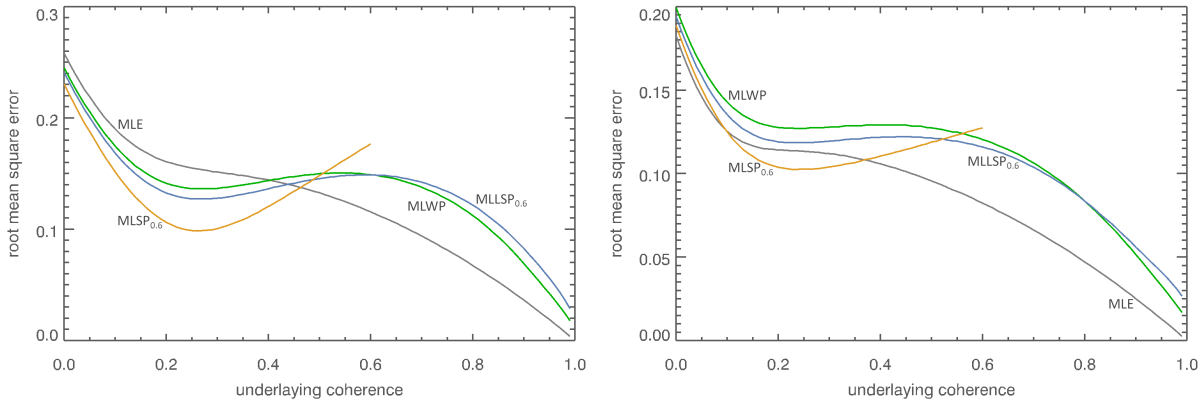
4.2.3 Characteristics for Large *N*

In the previous sections, the advantage of the ML methods is demonstrated for small sample sizes. The characterization for large sample sizes is provided for $N = 15$ in Fig. 4.10a and $N = 30$ in Fig. 4.10b. For $N = 15$, the RMSE of the ML methods perform still better compared to the sample estimator. The MLWP is more efficient for all underlying coherence magnitudes up to 0.40, the $\text{MLLSP}_{0.6}$ method up to 0.44 and the $\text{MLSP}_{0.6}$ estimator up to 0.46. However, the test case $N = 30$ demonstrates in Fig. 4.10b, only the $\text{MLSP}_{0.6}$ performs better than the sample estimator. Practically, the ML methods can be recommended for all sample sizes $N \leq 30$. It is noteworthy that the strict prior performs best also with large sample sizes. As a result, the ML methods improve the performance only for low coherences if the sample size is small.

4.2.4 Sentinel-1 Application Demonstration

Figure 4.11a visualizes the test case with 512×512 i.i.d. samples by the radar backscatter amplitude. The coherence magnitude from the sample estimator is visualized in Fig. 4.11b. Using identical estimation windows, the respective ML result is shown in Fig. 4.11d. In this example, the ML coherence magnitude is estimated locally adaptive with respect to the prior from MLWP, $\text{MLLSP}_{0.6}$, $\text{MLLSP}_{0.4}$ or $\text{MLSP}_{0.4}$. It can be seen that the estimation performance now depends not only on the window size but mainly on the prior and its strictness.

To give an intuitive idea of the effect of different priors and various parameters, Fig. 4.11c visualizes a composition of coherence estimates. In this figure from left to right, the result from the sample estimator, MLWP, $\text{MLLSP}_{0.6}$, $\text{MLLSP}_{0.4}$, $\text{MLSP}_{0.6}$, and $\text{MLSP}_{0.4}$ can be compared. Similar coherence magnitudes are observed for all but the last two columns. It



(a) $N = 15$. *(b)* $N = 30$.

Figure 4.10: RMSE characteristic of ML estimators for $N = 15$ (left) and $N = 30$ (right) comparing the sample estimator $\hat{\gamma}_\text{s}$ (gray), MLWP (green), $\text{MLLSP}_{0.6}$ (blue) and $\text{MLSP}_{0.6}$ (orange).

follows that the less strict prior can robustly cope with an underlying coherence greater than the prior parameter $\gamma_{\max}$.

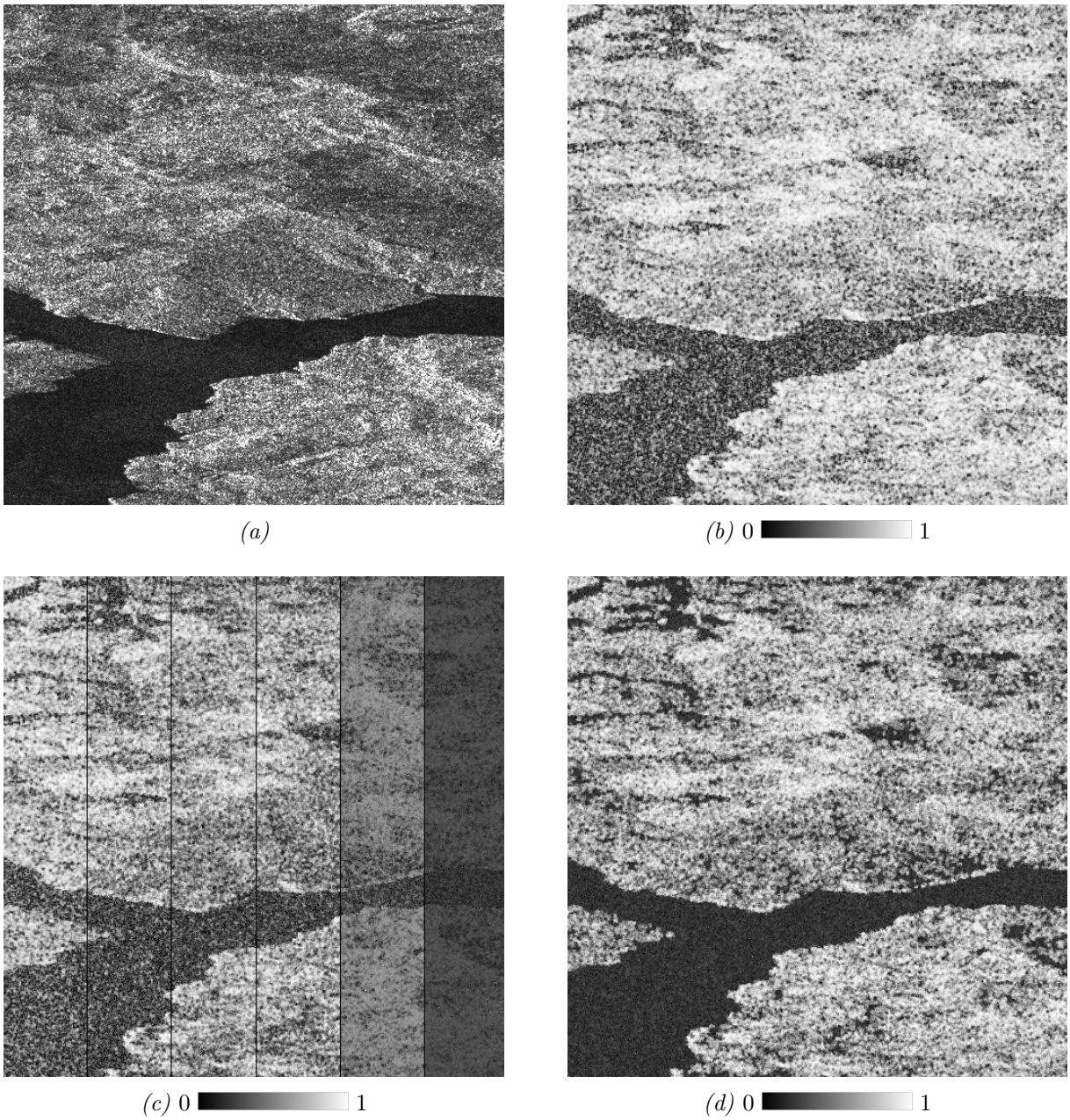


Figure 4.11: Sentinel-1 test case with $N = 9$ for machine learning estimators. (a) Radar backscatter amplitude. (b) Coherence magnitude from sample estimator $\hat{\gamma}_{\mathrm{s}}$. (c) Composite of coherence estimates from sample estimator, MLWP, $\mathrm{MLLSP}_{0.6}$, $\mathrm{MLLSP}_{0.4}$, $\mathrm{MLSP}_{0.6}$, $\mathrm{MLSP}_{0.4}$ from left to right. (d) ML coherence magnitude estimated locally adaptive from MLWP, $\mathrm{MLLSP}_{0.6}$, $\mathrm{MLLSP}_{0.4}$ or $\mathrm{MLSP}_{0.4}$.

4.3 Composite Estimation

4.3.1 Concept Testing

As a principle test, it was checked whether any one composite estimator could be better than available estimators, namely the sample estimator (2.11), the MLWP and the $\text{MLSP}_{0.6}$ from section 4.2. For this purpose, a composite estimator with many different estimators and subsample sizes was implemented, which corresponds to the configuration *CW_N30_G2G3G4G5G30W2W3W4W5*. Figure 4.12 shows that this estimator is better than the standard sample estimator and the MLWP and is even better than the $\text{MLSP}_{0.6}$. The latter is particularly noteworthy because the better RMSE was achieved without a prior.

4.3.2 Test of Two Partial Estimators Only

This experiment assesses whether so many partial estimators are actually required as in the principle test above with the configuration *CW_N30_G2G3G4G5G30W2W3W4W5*. It also answers the questions of whether a much simpler configuration namely *CW_N30_G2G30* can work better than the known estimators and whether it makes sense to use more partial estimates and thus accepting more computing and training effort. In section 4.2, the situation was that ML already outperforms the sample estimator. However, for large sample sizes N, the performance of ML is decreased and needs to be improved by prior knowledge, e.g., $\text{MLSP}_{0.6}$ in Fig. 4.10b in order to keep the advantage. Figure 4.13 shows that the composite estimator without prior *CW_N30_G2G30* is better than the sample estimator (*G30* in gray) and the $\text{MLSP}_{0.6}$ (blue) - even without prior. Clearly, the many estimators

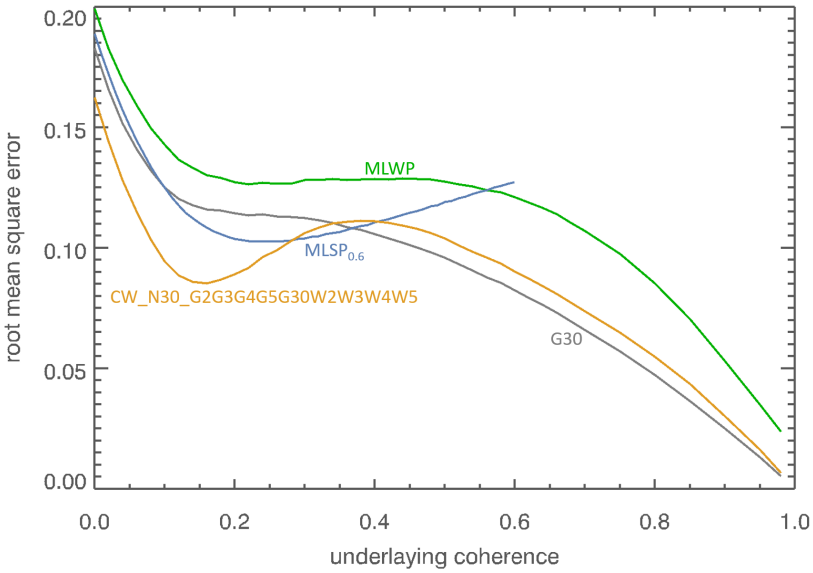


Figure 4.12: Principle test with oversized estimator with accepted redundancy.

in cw_n3_g2g3g4g5g3w2w3w4w5 from the principle test above are not even necessary. The experiment confirms the expectation that it is advantageous for a composite estimator if the statistical properties of the estimators used are as different as possible. This is the case if the number of subsamples is as different as possible. This experiment also demonstrates that the partial sample estimates ($G3$, $G4$, $G5$) and the machine learning estimates ($W2$, $W3$, $W4$ and $W5$) do not provide any additional features. In practice, the small subsample size of 2 is even very advantageous. It allows using all available samples if the given sample size N is divisible by two.

4.3.3 Testing the Need for Different Statistics

This test case answers the question of whether the difference in the statistics of the partial estimators, which is for example caused by the subsample size, is necessary. This is answered experimentally by comparing the configurations cw_n9_g2g9 and cw_n9_g2, i.e., the single very precise estimator $G30$ is left out. Figure 4.14 shows that there is a threshold above which cw_n9_g2 is better than the sample estimator $G9$ ($\gamma = 0.18$) and better than the composite estimator cw_n9_g2g9 ($\gamma = 0.35$). However, the configuration cw_n9_g2g9 outperforms the sample estimator using all $N = 9$ samples in a larger range, i.e., for all small coherences up to 0.53. It confirms the principle that different subsample sizes are to be preferred for composite estimators.

4.3.4 Testing the Need for an Intermediate Sample

The experiments above have demonstrated that the partial estimators should be as different as possible. This section checks whether an additional intermediate sample is

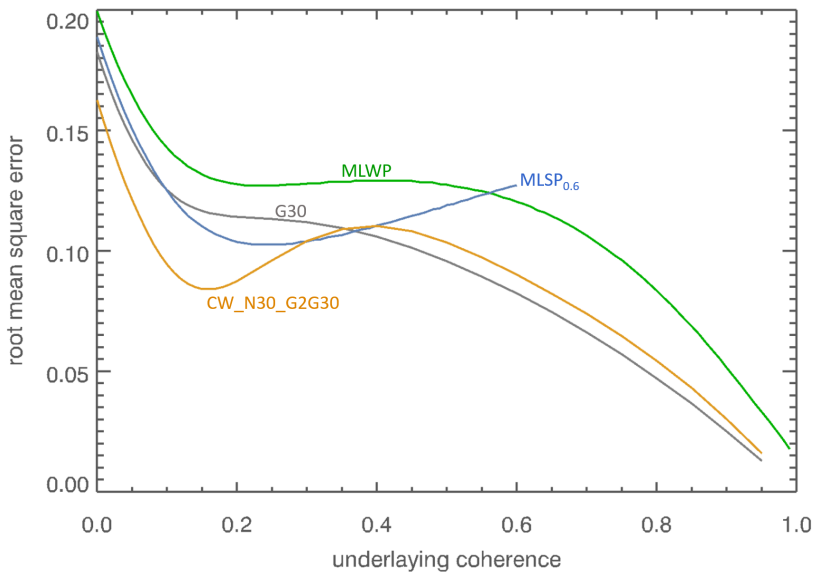


Figure 4.13: Comparison of minimal estimator cw_n30_g2g30 with known estimators.

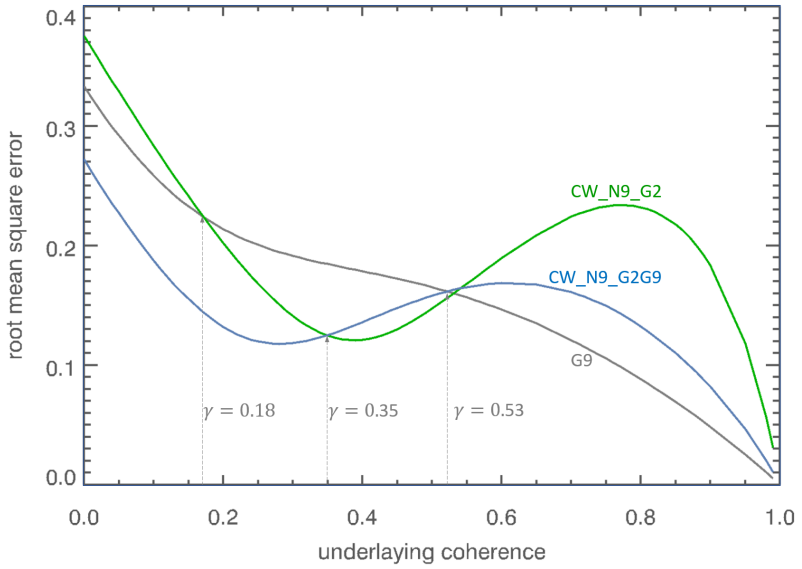


Figure 4.14: Comparison of *CW_N9_G2G9* and *CW_N9_G2* checking for the need of two different statistics.

worthwhile for large samples. This is an option that arises for very large samples. Practically, *CW_N60_G2G60* and *CW_N60_G2G30G60* are compared. Figure 4.15 shows that there is no advantage if another estimator (*G30* in this example) is added.

4.3.5 Comparison of Different Subsample Sizes

This test case examines whether it makes a difference to choose a subsample size $S = 2$ or $S = 3$. Figure 4.16 shows that it does not matter whether subsamples are formed for the estimators *G2* or *G3* if the number of given samples is a divisor of 2 and 3. Finally, the composite estimators *CW_N9_G2G9* and *CW_N9_G3G9* are equally good. We can therefore choose the one that best partitions the total samples, e.g., $N = 30$ with 15×2 samples however $N = 9$ with 3×3 samples. Every sample is then used and the number of features to learn from is smaller.

4.3.6 Comparison of Strict Prior Test Cases

Now, it is assessed whether prior information can also be used in this method in accordance with the principle from section 4.3. The strict prior is used as a test case. This means that this prior is used for the training of the composite estimator itself and one or all partial estimators restrict their estimation onto this prior range. I.e., the composite estimator inherits the constraints of the partial estimator, otherwise there is a risk that the partitioned input data do not correspond to their prior. Figure 4.17 compares the estimators *CS(0.6)_N30_S(0.6)2G30*, *CS(0.6)_N30_G2S(0.6)30*, and *CS(0.6)_N30_S(0.6)2S(0.6)30*, i.e., in the first, the frequent imprecise estimate *S(0.6)2* and in the second, the precise one-time

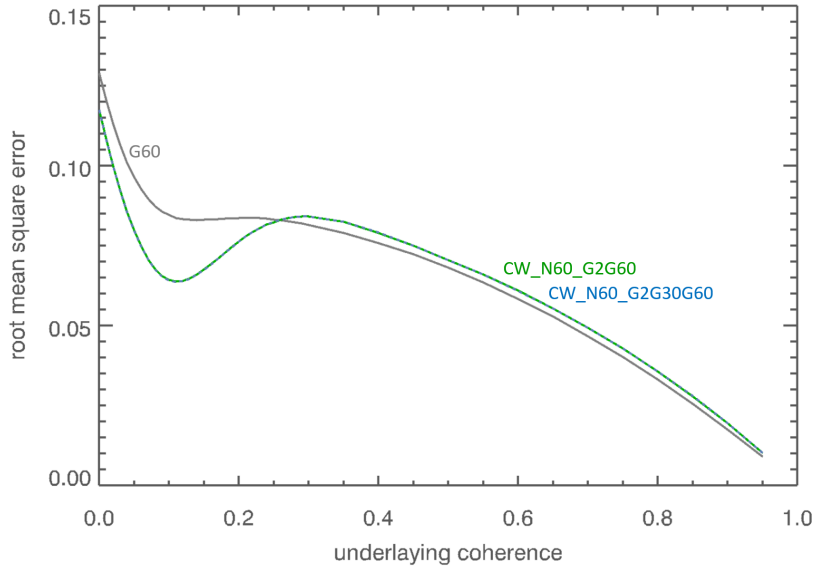


Figure 4.15: Comparison of CW_N60_G2G60 and $CW\_N60\_G2G30G60$ checking the the need for an intermediate sample.

estimate $S(0.6)30$ is improved by the prior. The third estimator restricts the estimation of all partial estimators into the prior range. Figure 4.17 shows that the performance is improved best in the first configuration $CS(0.6)\_N30\_S(0.6)2G30$.

4.3.7 Characteristics for $N = 9$ and $N = 30$

The estimator characteristics are assessed with test cases $N = 9$ and $N = 30$. Figure 4.18 compares the RMSE for the sample coherence estimator and the composite estimator with various variants.

The left column of the figure shows the results for $N = 9$. For small coherences, i.e., $\gamma < 0.53$, all composite estimators outperform the sample estimator. Clearly, the stricter the prior, the better is the RMSE of the estimate. The bias is compared in Fig. 4.18a. For small coherences, it is reduced in absolute value for all composite estimators. However, the absolute bias value is larger compared to the sample estimator for a higher underlying coherence. This figure also shows the property that, the stricter the prior, the better is the bias of the estimate. Figure 4.18c shows that the prior significantly improves the standard deviation of the estimation.

For $N = 30$, the RMSE is compared for the sample coherence estimator and different versions of the composite estimator in Fig. 4.18f. For small coherences, i.e., $\gamma < 0.36$, all composite estimators outperform the sample estimator. However, the less strict prior is no longer as effective. This means that the improvement in RMSE for $CL(0.4)\_N30\_L(0.4)2G30$ (orange graph) is insignificant compared to the estimator $CL(0.6)\_N30\_L(0.6)2G30$ (blue graph). In addition, the composite estimator without prior CW_N30_G2G30 (green graph) is even similar to a less strict prior with $\gamma_{\max} = 0.6$ in terms of performance. In contrast, the

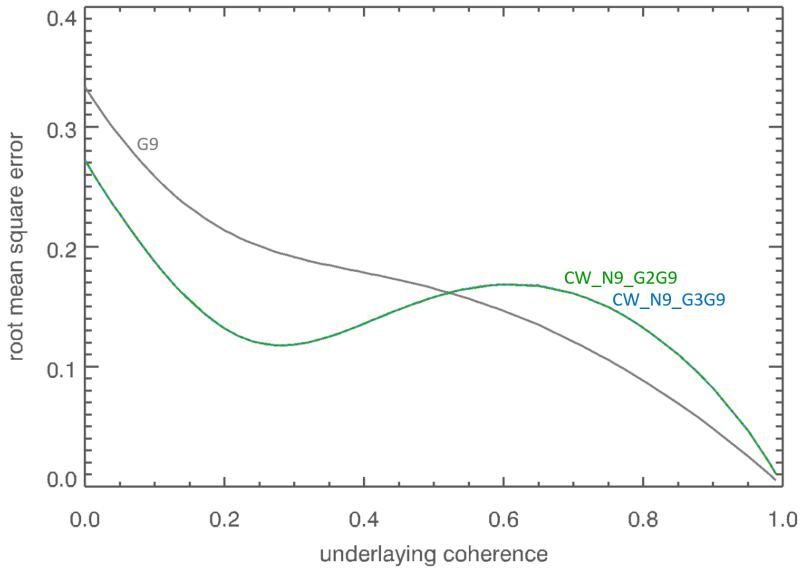


Figure 4.16: Comparison of *CW_N9_G2G9* and *CW_N9_G3G9* checking for different subsample sizes.

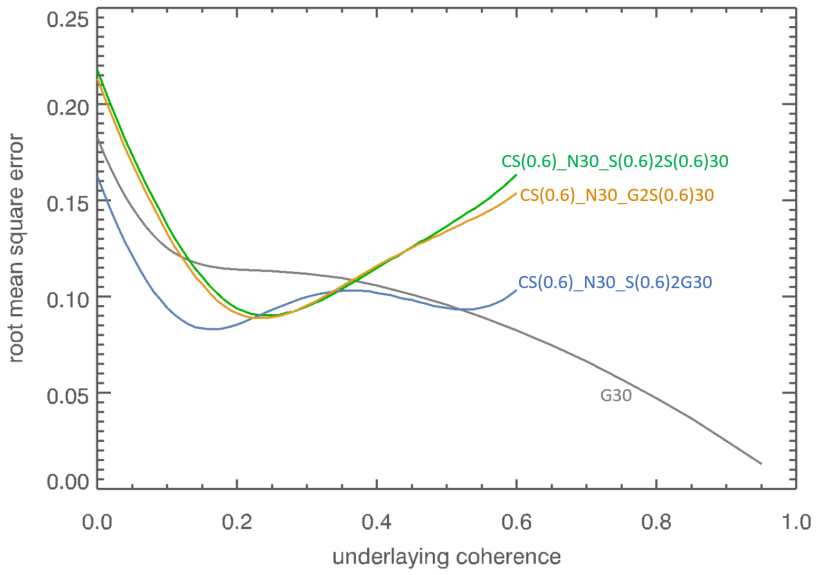


Figure 4.17: Comparison of possible configurations to process strict prior, i.e., *CS*(0.6)*_N*30*_S*(0.6)2*G*30, *CS*(0.6)*_N*30*_G*2*S*(0.6)30 and *CS*(0.6)*_N*30*_S*(0.6)2*S*(0.6)30.

strict prior composite estimator continues to be improving the RMSE. Figure 4.18b showing the bias and Fig. 4.18d with the estimation standard deviation confirm the usefulness of the strict prior with the test cases *CS*(0.6)_*N*30_*S*(0.6)2*G*30 (light green graph) and *CS*(0.6)_*N*30_*S*(0.6)2*G*30 (red graph).

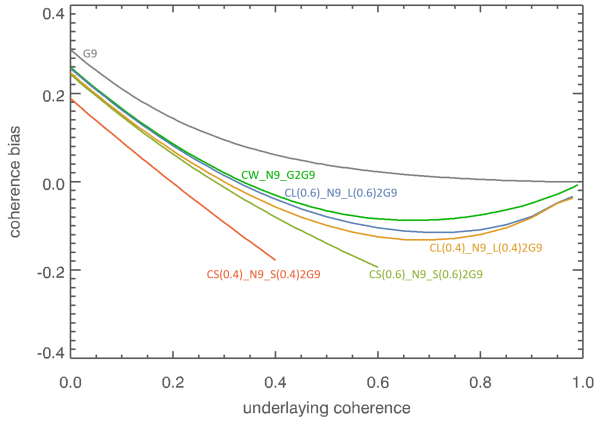
4.3.8 Characteristics for Large *N*

Next, the influence of the available sample size N on the estimation performance is discussed and four characteristics are illustrated. Figure 4.19 compares similar constructed composite estimators without and with strict prior for sample sizes $N = 9$, $N = 30$, $N = 100$ and $N = 200$. First, the best improvements are achieved for small samples. Quantitatively, this is described below for the most difficult case, i.e., coherence $\gamma = 0$.

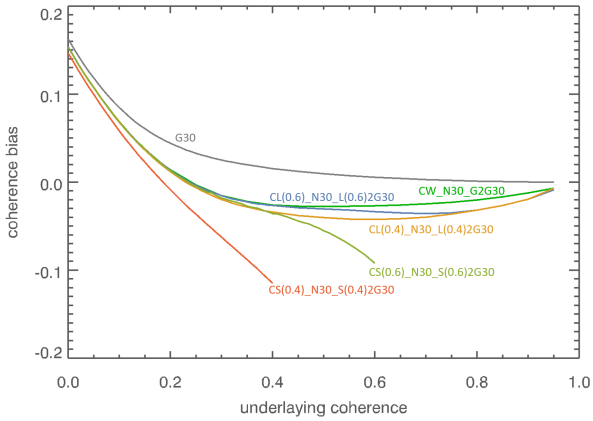
Figure 4.20a shows that a relative improvement of 30.6 % is achieved for $N = 3$, even without a prior. With $N = 200$ samples there is still an improvement of 7.4 % as Fig. 4.19d and 4.20a show. Second, for each sample size there is a specific coherence at which the improvement is maximum. For example, the smallest RMSE is achieved at $\gamma = 0.28$ with $N = 9$ and $\gamma = 0.15$ with $N = 30$ for the composite estimator without prior (green graphs). For the composite estimator with strict prior (blue graph), there exist two areas with better performance for large samples as demonstrated in Fig. 4.19c and 4.19d. Third, Fig. 4.19 also demonstrates that the coherence range for which a better performance of the composite estimator without prior (green graph) is achieved compared to the sample estimator (gray graph) becomes smaller with increasing sample size. Figure 4.20b provides the experimentally evaluated relationship. Fourth, increasing the sample size, all priors act less as demonstrated in 4.19c and 4.19d for the strict prior. Both figures show that the RMSE of the estimator with strict prior becomes more and more similar to the RMSE of the estimator without prior as the sample size increases.

4.3.9 Sentinel-1 Application Demonstration

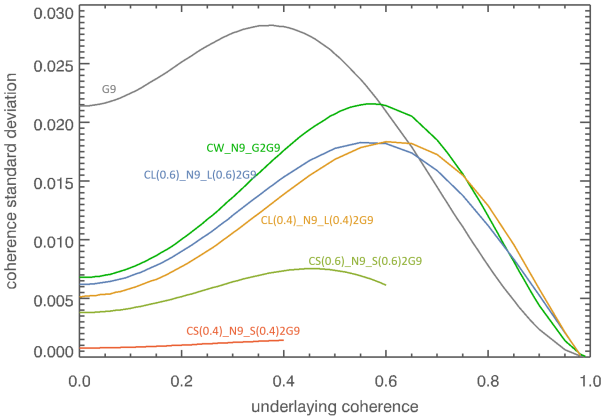
Figure 4.21a visualizes the test case with 512×512 samples by the radar backscatter amplitude. The coherence magnitude from the conventional sample estimator is visualized in Fig. 4.21b. It is universal because it does not require prior information. The composite estimator *CW*_*N*9_*G*2*G*9 has the same general applicability since it also does not require a prior. As Fig. 4.21c indicates, it is the better universal estimator compared to Fig. 4.21b. Figure 4.21d illustrates the estimation performance when prior information is included for each pixel and the estimators *CW*_*N*9_*G*2*G*9, *CL*(0.6)_*N*9_*L*(0.6)2*G*9, *CL*(0.4)_*N*9_*L*(0.4)2*G*9 or *CS*(0.4)_*N*9_*S*(0.4)2*G*9 are applied locally adaptively.



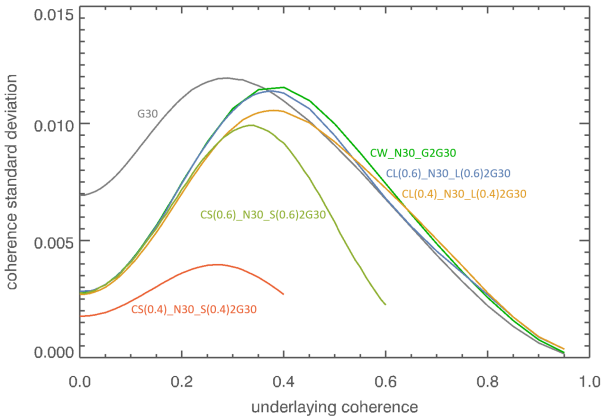
(a) Bias $\gamma_*^{(\text{bias})}$ for $N = 9$.



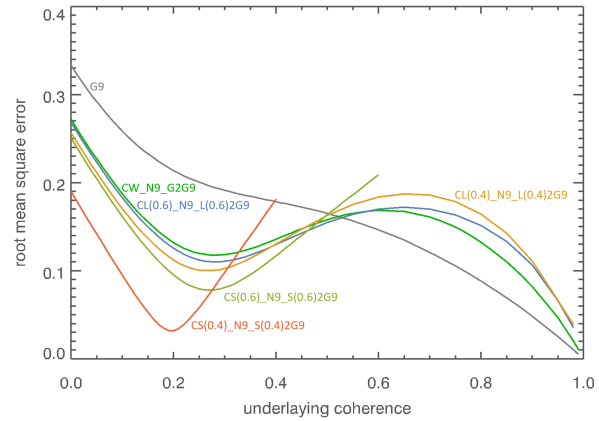
(b) Bias $\gamma_*^{(\text{bias})}$ for $N = 30$.



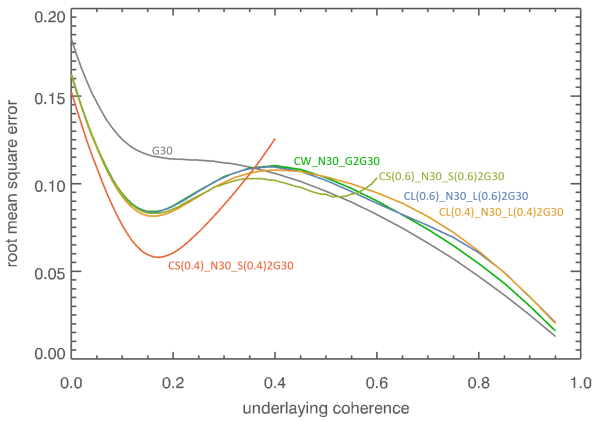
(c) Standard deviation $\gamma_*^{(\sigma)}$ for $N = 9$.



(d) Standard deviation $\gamma_*^{(\sigma)}$ for $N = 30$.



(e) RMSE $\gamma_*^{(\text{RMSE})}$ for $N = 9$.



(f) RMSE $\gamma_*^{(\text{RMSE})}$ for $N = 30$.

Figure 4.18: Characteristic of composite estimators for $N = 9$ left column and $N = 30$ right column.

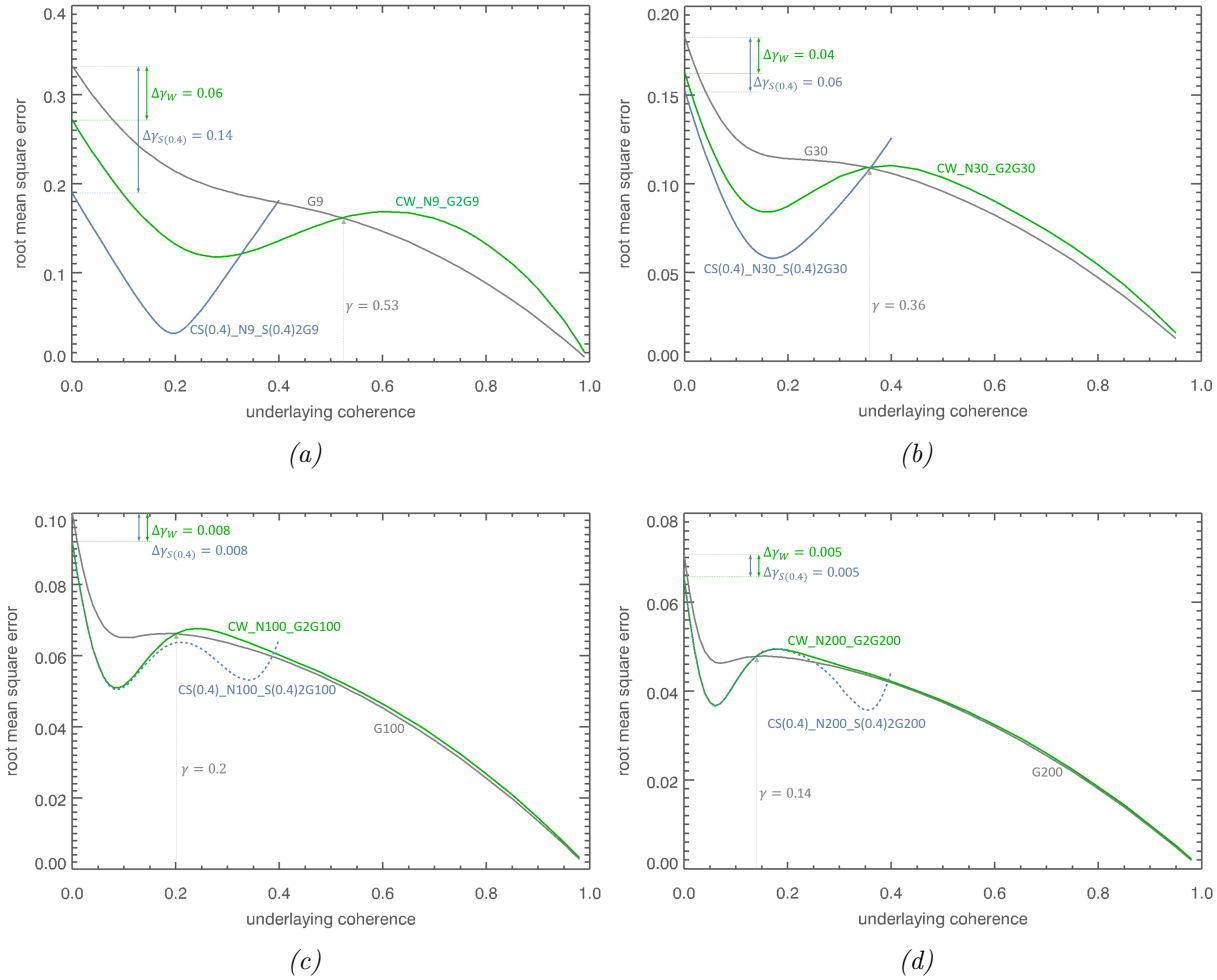
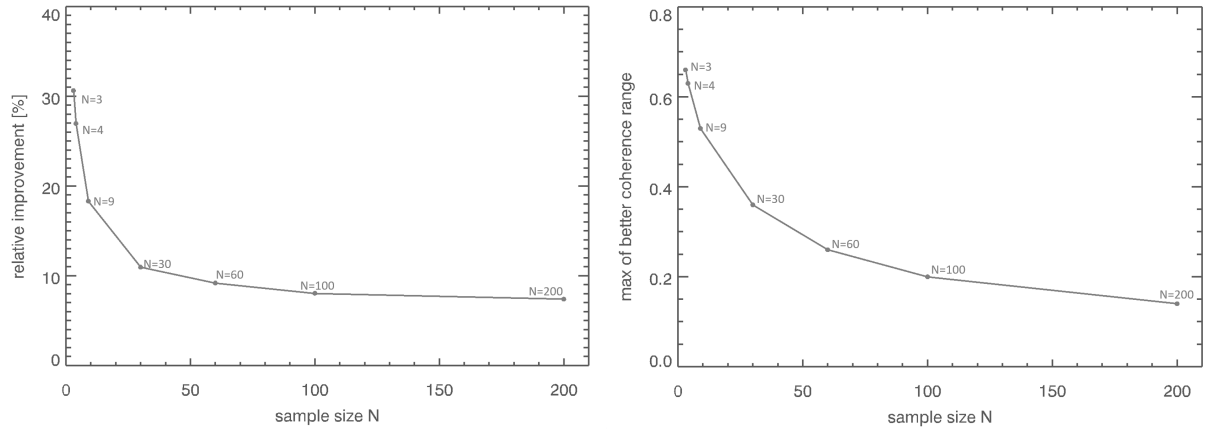


Figure 4.19: Comparison of sample size a) $N = 9$, b) $N = 30$, c) $N = 100$ and d) $N = 200$ checking the effect on the RMSE performance.



(a) Relative improvement [%] of composite estimators without prior with respect to the sample estimator at $\gamma = 0$.

(b) Maximum coherence up to which the composite estimator without prior is better than the sample estimator.

Figure 4.20: Characteristic of composite estimators with respect to the sample size N.

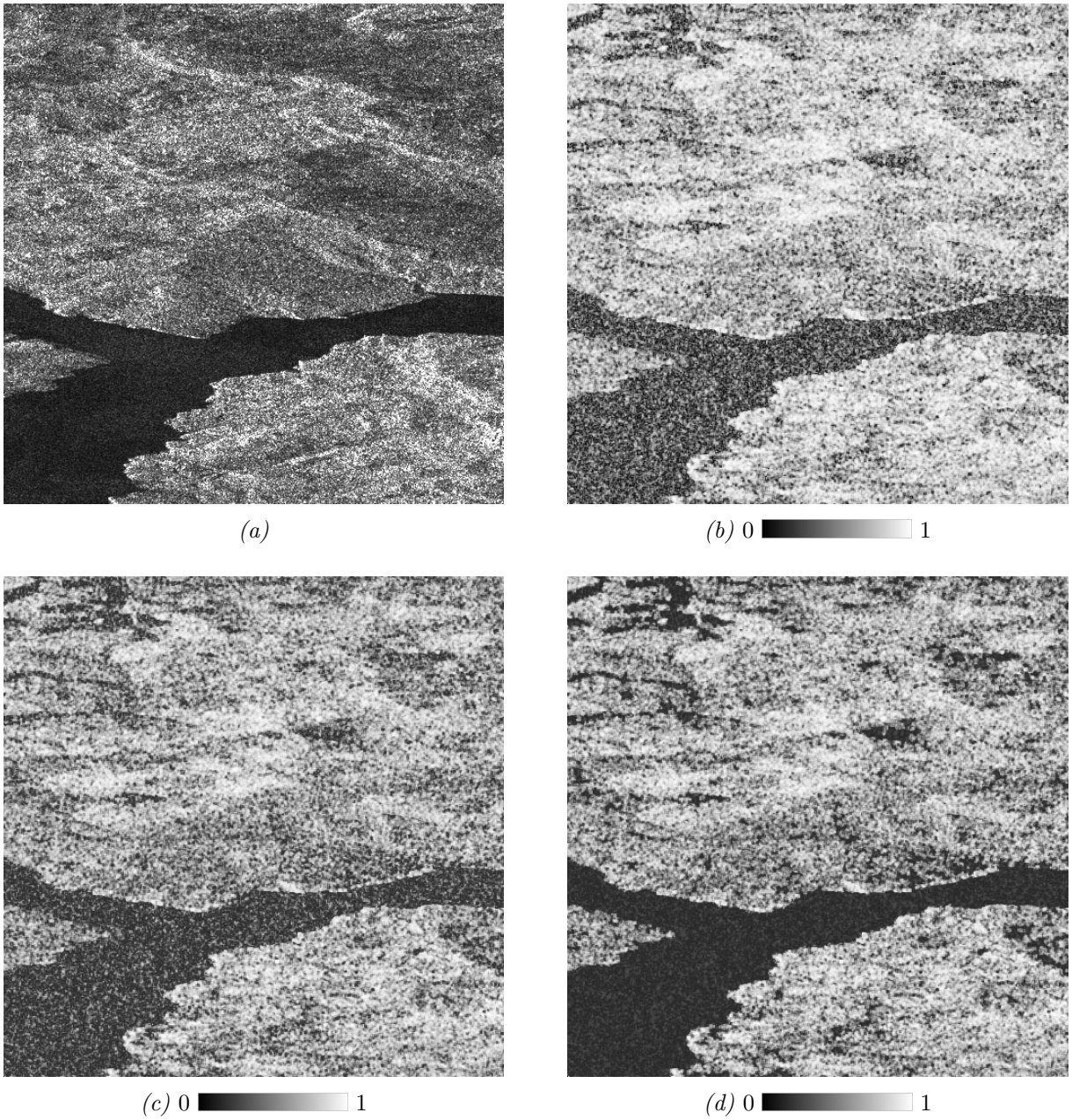


Figure 4.21: Sentinel-1 test case with $N = 9$ for composite estimators. (a) Radar backscatter amplitude. (b) Coherence magnitude from sample estimator (2.11). (c) Coherence magnitude from composite estimator *CW_N*9*_G*2*G*9, i.e., without prior. (d) Coherence magnitude estimated from composite estimators locally adaptive from configurations *CW_N*9*_G*2*G*9, *CL*(0.6)*_N*9*_L*(0.6)2*G*9, *CL*(0.4)*_N*9*_L*(0.4)2*G*9 or *CS*(0.4)*_N*9*_S*(0.4)2*G*9.

5 Discussion

5.1 On the CCG Signal Model and Simulation

In this thesis, three new methods for estimating the coherence magnitude of jointly CCG signals are developed. This signal model is not a limitation on applicability because the signal characteristic is fundamental in SAR and is accepted for DSs observed with medium resolution SAR and proven by many InSAR applications. For example, it is used by Touzi & Lopes (1996) to derive the coherence estimation characteristics (2.12), (2.17), (2.18) and by Bamler (2000) to predict the coregistration accuracy (2.32). Furthermore, the performance analysis, e.g., by (2.28), (2.29), and optimization of successful SAR missions, e.g., SRTM (Breit & Bamler 1998) and TanDEM (Krieger et al. 2005) as well as the phase-triangulation algorithms (Rocca 2007, Ferretti et al. 2011, Fornaro et al. 2015, Cao et al. 2015) are based on this specific signal model by (2.21).

Although the ML procedures and the characterizations of the developed coherence estimators are based on extensive simulations, the results are applicable to real data. This is due to the fact that the simulation accurately generates actual CCG processes without any approximations, assumptions, or models. Obviously, there is no simulation or assumption about the components of the InSAR signal

$$\varphi_i = \varphi_i^{(\mathrm{h})} + \varphi_i^{(\mathrm{defo})} + \varphi_i^{(\mathrm{atmo})} + \varphi_i^{(\mathrm{noise})}, \tag{5.1}$$

which are the topography phase $\varphi_i^{(\mathrm{h})}$, the displacement phase $\varphi_i^{(\mathrm{defo})}$, the atmosphere effect $\varphi_i^{(\mathrm{atmo})}$ and the noise $\varphi_i^{(\mathrm{noise})}$. The most general and realistic use case is that φ_i is uniformly distributed, i.e., $\varphi_i \sim U(-\pi, \pi)$. Indeed, a phase simulation of the InSAR phase components would increase the dimensionality of the problem and make simulation and ML training more difficult. Because the simulation generates realistic data, the Sentinel-1 test cases in sections 4.1.5, 4.2.4 and 4.3.9 work so well. The only requirement for the ML is to provide sufficient training data from simulations.

Nevertheless, the developed algorithms are only valid for CCG SHPs. However, this is a very general and natural class of signals and also means that the procedures are not restricted to InSAR but can generally be applied to coherence estimation problems from joint CCG data.

In practice, SAR observes not only DSs described in section 2.1.1 but also other surfaces and scatterers, e.g., the overlay of DSs, a single dominant PS and multiple PSs in a resolution cell. Karakus et al. (2021) give a variety of statistical models for SAR data in Table 1 for cases where the CCG principle does not apply. In principle, these statistical models

describe heterogeneous pixel areas that are not SHP, or in other words, do not consist of i.i.d. samples. In practice, heterogeneous pixel areas can only occur in the coherence estimation if the SHP detection was incorrect. Therefore, to avoid error propagation, a robust coherence estimation procedure is desirable however out of scope in this thesis. For non i.i.d. signals from DSs, e.g., the overlay of DSs, or signals from PSs, the conventional estimator is also inappropriate because the prerequisites $\mathrm{E}\{\boldsymbol{X}_1\} = 0$ and $\mathrm{E}\{\boldsymbol{X}_2\} = 0$ are not met in (2.8). It is also not surprising that the characterization of the sample coherence cannot be transferred to heterogeneous signals or signals with outliers.

5.2 On the Estimator Characteristic

Compared to the sample estimator, the developed estimators improve the estimation of small coherences. And, they better estimate the coherence from a small sample size. Another characteristic is that the smaller the sample size, the better the estimators improve the estimation. These characteristics are expected and confirmed by the theory. Efficiency is the criterion for the best possible estimator. An estimator is efficient if it achieves the CRLB or the best possible MSE or RMSE. The latter two are determined by the bias and the variance. Two properties of the sample estimator make it, for large N and coherence magnitude close to 1, the efficient estimator that cannot be outperformed. Firstly, the sample estimator is asymptotically unbiased, i.e., for large N and/or high coherence γ. And secondly, the MLEs asymptotically have minimal variance. Also, the sample estimator is consistent, meaning that it estimates more accurately as the sample size increases. Of course, both cases of $\gamma = 1$ and $N = \infty$ are rather theoretical. In practice, however, these facts mean that other estimators cannot be better for these extreme configurations, and it becomes more difficult for other estimators the closer one gets to these extremes. I.e., N is large and/or the coherence is large. This thesis has expanded the range of coherence and sample size at which the developed estimators outperform the sample estimator, starting with the empirical Bayes estimators, the ML based estimations, and finally the composite estimators.

6 Conclusion and Outlook

6.1 Conclusion

In this thesis, three new methods for estimating the coherence magnitude of jointly CCG signals are developed. This signal characteristic is accepted for distributed scatterer observed with medium resolution SAR. First, the empirical Bayesian method estimates from the statistic of the posterior distribution. Three statistics are tested, namely MAP, EAP and MEDAP. Concerning the bias, the simulation shows that EAP is better than MEDAP and the latter is better than MAP. Second, an improvement in terms of estimation accuracy and computational complexity is the ML coherence estimator. Third, a further improvement in terms of estimation accuracy is the estimator based on composition of coherence subsample estimates and ML. It extends and uses also the straight forward ML estimator. In practice, both ML methods can be implemented with available ML libraries and are independent of which method is used, such as a decision tree, a random forest or a neural network.

The following characteristics of the estimators can be concluded:

- The developed estimators offer two main advantages compared to the conventional sample estimator. All improve the estimation of small coherences. And, they better estimate the coherence from small sample sizes. All estimators are better than the sample estimator in their specific coherence range, which depends on the number of InSAR samples N, the underlying coherence γ and the prior.
- Another characteristic is that the smaller the sample size, the better the developed estimators improve the estimation.
- As the sample size increases, the bias vanishes and the estimators are asymptotically unbiased.
- The principle of empirical Bayes improves the coherence estimation for a sample size $2 \leq N \leq 15$ compared to the conventional sample estimator.
- The ML based estimation is advantageous for $2 \leq N < 30$ and for a larger sample size, prior information needs to support the estimation for an advantage over the sample estimator.
- For the composite estimator, the superior performance for large samples compared to the sample estimator is demonstrated for the test cases $N = 100$ and $N = 200$. Surprisingly, already 3 InSAR samples are sufficient for an improved estimation.

- All developed estimators support prior information in form of a prior PDF $p(\gamma)$.

In principle, all estimators are suitable for typical InSAR scenarios. First, the uninformative prior improves the estimation without prior knowledge and is generally applicable. Second, the informative priors include an assumption on the maximum coherence of the underlying coherence. Indeed, such information is available in InSAR based on time series. The strict prior limits the estimates inside the assumed range and the less strict prior favors estimates in this range. The more information is used and the stricter the prior, the more accurate the estimate will be. The strict prior achieves the best performance with respect to the RMSE. The informative priors are described by a single parameter only, i.e., the maximum underlying coherence $\gamma_{\max}$.

The performance and advantages are the reasons why the composite estimator is suitable and recommended for implementation in operational InSAR systems. It supports small and large sample sizes and has the best estimation performance compared to the other methods. An advantage worth emphasizing is its estimation performance without prior. This makes the estimator universally applicable and comparable with the conventional sample estimator. However, it requires to also implement the direct ML estimators.

The developed estimators are not limited to InSAR, however can be generally applied to coherence estimation problems from joint CCG processes.

6.2 Outlook

There are several ways to advance the research conducted in this thesis, and the following subproblems offer one way to do so.

Improve Prior: In this thesis, the prior is straight forward and parameterized by a single parameter. It is to be expected that a more complicated prior, e.g., with more parameters and therefore a more precise prior, will lead to more accurate estimates. Furthermore, additional prior information can be added. One example is the assumption of the same intensity in both channels.

Improve Algorithms: The developed coherence estimators start from i.i.d. samples, i.e., SHP. The detection of these areas is not trivial, is still an active field of research and can be erroneous. To avoid error propagation, a robust coherence estimation procedure is desirable. The algorithms should be gradually improved to identify and tolerate first one, then two and finally up to 50 % outliers (i.e. a mixture of SHP or additional PSs).

Another useful extension together with the robust estimation is a recursive estimation, in which a given SHP is extended by an InSAR sample and the intermediate results of the previous estimation are used advantageously. This further development can be used for SHP detection.

The estimation precision of an executed coherence estimation is the essential parameter for the evaluation of an error propagation in an application. Estimating the estimation

precision is therefore desirable and would also be useful with the current sample estimator, however is typically not even implemented at present.

Operational Setup: Depending on the ML library or framework used, the computational effort for the actual estimation process differs. Fast and uncomplicated execution is desirable. A complexity and performance analysis provide decision criteria for the selection of a framework or software library that is particularly suitable for operational systems.

Many estimation functions need to be trained for the ML methods, i.e., one for each prior and prior parameter. This is a considerable effort in terms of the required computing performance. The computational effort is also high, as a lot of data samples have to be simulated for the training. The more data is simulated, the better will be the implemented estimator. This effort should be bundled and a free library of coherence estimators should be created and shared.

Demonstrate Applications: The developed estimators are not limited to InSAR, however can be generally applied to coherence estimation problems from joint CCG data. It is desirable to demonstrate such use cases in different work domains that differ from InSAR.

The composite estimation approach using ML provides a generally usable principle for parameter estimation. It is worth to test and demonstrate any omnipresent parameter estimation, e.g., variance, amplitude or intensity from real samples. An obvious application for SAR data would be incoherent change detection. However, it is equally desirable to demonstrate such use cases in work domains that differ from SAR and InSAR.

Bibliography

Abdelfattah, R. & Nicolas, J.-M. (2006), 'Interferometric SAR coherence magnitude estimation using second kind statistics', *IEEE Transactions on Geoscience and Remote Sensing* **44**(7), 1942–1953.

Adam, N. (2022), 'Empirical Bayesian estimation of the interferometric SAR coherence magnitude', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **15**, 6306–6323.

Adam, N. (2023), 'Interferometric SAR coherence magnitude estimation by machine learning', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **16**, 3034–3044.

Adam, N. (2024), 'SAR coherence estimation by composition of subsample estimates and machine learning', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **17**, 15957–15970.

Adam, N. A., Kampes, B., Eineder, M., Worawattanamateekul, J. & Kircher, M. (2003), The development of a scientific permanent scatterer system, *in* K. Jacobsen & C. Heipke, eds, 'Proceedings of the Joint ISPRS/EARSeL Workshop "High Resolution Mapping from Space 2003"', ISPRS.

Ansari, H., De Zan, F. & Bamler, R. (2017), 'Sequential estimator: Toward efficient InSAR time series analysis', *IEEE Transactions on Geoscience and Remote Sensing* **55**(10), 5637–5652.

Bamler, R. (2000), Interferometric stereo radargrammetry: absolute height determination from ERS-ENVISAT interferograms, *in* 'Proc. of IEEE International Geoscience and Remote Sensing Symposium', Vol. 2, pp. 742–745.

Bamler, R. & Hartl, P. (1998), 'Synthetic aperture radar interferometry', *Inverse Problems* **14**(R1).

Berardino, P., Fornaro, G., Lanari, R. & Sansosti, E. (2002), 'A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms', *IEEE Transactions on Geoscience and Remote Sensing* **40**(11), 2375–2383.

Berger, J. (1985), *Statistical Decision Theory and Bayesian Analysis*, Springer Series in Statistics, Springer.

Bourbigot, M. (2016), 'Sentinel-1 product definition', ESA-TN. S1-RS-MDA-52-7440.

Breit, H. & Bamler, R. (1998), An InSAR processor for on-board performance monitoring of the SRTM/X-SAR interferometer, *in* 'Proc. of IEEE International Geoscience and Remote Sensing Symposium', Vol. 4, pp. 1904–1907.

Breit, H., Fritz, T., Balss, U., Niedermeier, A., Lachaise, M. & Schättler, B. (2008), TerraSAR-X SAR data processing, *in* 'CEOS SAR Calibration and Validation Workshop 2008'.

Cao, N., Lee, H. & Jung, H. C. (2015), 'Mathematical framework for phase-triangulation algorithms in distributed-scatterer interferometry', *IEEE Geoscience and Remote Sensing Letters* **12**(9), 1838–1842.

Carter, G., Knapp, C. & Nuttall, A. (1973), 'Statistics of the estimate of the magnitude-coherence function', *IEEE Transactions on Audio and Electroacoustics* **21**(4), 388–389.

Chen, T., Benesty, M., Tang, Y., Zhu, N., Yuan, J., Cho, H., Mitchell, R., Liu, H., He, T., Khotilovich, V., Xu, B., Lebedev, S., Lundberg, S. & Smirnov, E. (2022), 'XGBoost eXtreme Gradient Boosting library'.
URL: *https://github.com/dmlc/xgboost*

Chen, T. & Guestrin, C. (2016), XGBoost: A scalable tree boosting system, *in* 'Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining', KDD '16, ACM, New York, NY, USA, pp. 785–794.

Colesanti, C., Ferretti, A., Novali, F., Prati, C. & Rocca, F. (2003), 'SAR monitoring of progressive and seasonal ground deformation using the permanent scatterers technique', *IEEE Transactions on Geoscience and Remote Sensing* **41**(7), 1685–1701.

Costantini, M., Minati, F., Trillo, F., Ferretti, A., Passera, E., Rucci, A., Dehls, J., Larsen, Y., Marinkovic, P., Eineder, M., Brcic, R., Siegmund, R., Kotzerke, P., Kenyeres, A., Costantini, V., Proietti, S., Solari, L. & Andersen, H. S. (2022), EGMS: Europe-wide ground motion monitoring based on full resolution InSAR processing of all Sentinel-1 acquisitions, *in* 'Proc. of IEEE International Geoscience and Remote Sensing Symposium', pp. 5093–5096.

Crosetto, M., Solari, L., Mróz, M., Balasis-Levinsen, J., Casagli, N., Frei, M., Oyen, A., Moldestad, D. A., Bateson, L., Guerrieri, L., Comerci, V. & Andersen, H. S. (2020), 'The evolution of wide-area DInSAR: From regional and national services to the European Ground Motion Service', *Remote Sensing* **12**(12). paper 2043.

Derauw, D. (1995), Phase unwrapping using coherence measurements, *in* G. Franceschetti, C. J. Oliver, J. C. Shiue & S. Tajbakhsh, eds, 'Synthetic Aperture Radar and Passive Microwave Sensing', Vol. 2584, International Society for Optics and Photonics, SPIE, pp. 319–324.

Eineder, M., Breit, H., Adam, N., Holzner, J., Suchandt, S. & Rabus, B. (2001), SRTM X-SAR calibration results, *in* 'Proc. of IEEE International Geoscience and Remote Sensing Symposium', Vol. 2, pp. 748–750.

Even, M., Westerhaus, M. & Kutterer, H. (2024), 'German and European Ground Motion Service: a comparison', *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science* pp. 2512–2819.

Ferretti, A., Fumagalli, A., Novali, F., Prati, C., Rocca, F. & Rucci, A. (2011), 'A new algorithm for processing interferometric data-stacks: SqueeSAR', *IEEE Transactions on Geoscience and Remote Sensing* **49**(9), 3460–3470.

Ferretti, A., Prati, C. & Rocca, F. (1999), Non-uniform motion monitoring using the permanent scatterers technique, *in* 'Proceedings FRINGE99', ESA, pp. 431–436.

Ferretti, A., Prati, C. & Rocca, F. (2001), 'Permanent scatterers in SAR interferometry', *IEEE Transactions on Geoscience and Remote Sensing* **39**(1), 8–20.

Fornaro, G., Serafino, F. & Soldovieri, F. (2003), 'Three-dimensional focusing with multipass SAR data', *IEEE Transactions on Geoscience and Remote Sensing* **41**(3), 507–517.

Fornaro, G., Verde, S., Reale, D. & Pauciullo, A. (2015), 'CAESAR: An approach based on covariance matrix decomposition to improve multibaseline–multitemporal interferometric SAR processing', *IEEE Transactions on Geoscience and Remote Sensing* **53**(4), 2050–2065.

Freeman, A. (1992), 'SAR calibration: an overview', *IEEE Transactions on Geoscience and Remote Sensing* **30**(6), 1107–1121.

Frei, M. (2017), White paper: European Ground Motion Service (EU-GMS) - A proposed Copernicus service element, pp. 1–24.

Freund, Y. & Schapire, R. E. (1996), Experiments with a new boosting algorithm, *in* 'Proceedings of the Thirteenth International Conference on International Conference on Machine Learning', ICML'96, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, p. 148–156.

Gabriel, A. K. & Goldstein, R. M. (1988), 'Crossed orbit interferometry: theory and experimental results from SIR-B', *International Journal of Remote Sensing* **9**(5), 857–872.

Goldstein, R. M., Caro, E. R. & Wu, C. (1985), 'Method and apparatus for contour mapping using synthetic aperture radar'.

Goldstein, R. M., Engelhardt, H., Kamb, B. & Frolich, R. M. (1993), 'Satellite radar interferometry for monitoring ice sheet motion: Application to an antarctic ice stream', *Science* **262**(5139), 1525–1530.

Goodman, J. (2007), *Speckle Phenomena in Optics: Theory and Applications*, Roberts & Company.

Goodman, J. W. (1975), *Statistical Properties of Laser Speckle Patterns*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 9–75.

Goodman, J. W. (1976), 'Some fundamental properties of speckle∗', *J. Opt. Soc. Am.* **66**(11), 1145–1150.

Graham, L. (1974), 'Synthetic interferometer radar for topographic mapping', *Proceedings of the IEEE* **62**(6), 763–768.

Guarnieri, A. M. & Tebaldini, S. (2008), 'On the exploitation of target statistics for SAR interferometry applications', *IEEE Transactions on Geoscience and Remote Sensing* **46**(11), 3436–3443.

Hanssen, R. & Bamler, R. (1999), 'Evaluation of interpolation kernels for SAR interferometry', *IEEE Transactions on Geoscience and Remote Sensing* **37**(1), 318–321.

Hensley, S., Rosen, P. & Gurrola, E. (2000), The SRTM topographic mapping processor, *in* 'Proc. of IEEE International Geoscience and Remote Sensing Symposium', Vol. 3, pp. 1168–1170.

Ho, T. K. (1995), Random decision forests, *in* 'Proceedings of 3rd International Conference on Document Analysis and Recognition', Vol. 1, pp. 278–282.

Hooper, A., Zebker, H., Segall, P. & Kampes, B. (2004), 'A new method for measuring deformation on volcanoes and other natural terrains using InSAR persistent scatterers', *Geophyiscal Research Letters* **31**, 1–5.

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2021), *Linear Model Selection and Regularization*, Springer US, New York, NY, pp. 15–57.

Jiang, M., Ding, X. & Li, Z. (2014), 'Hybrid approach for unbiased coherence estimation for multitemporal InSAR', *IEEE Transactions on Geoscience and Remote Sensing* **52**(5), 2459–2473.

Just, D. & Bamler, R. (1994), 'Phase statistics of interferograms with applications to synthetic aperture radar', *Appl. Opt.* **33**(20), 4361–4368.

Kalia, A. C. (2017), User driven products in the context of the Ground Motion Service Germany, *in* 'Proc. of IEEE International Geoscience and Remote Sensing Symposium', pp. 1688–1691.

Kalia, A. C., Frei, M. & Lege, T. (2018), Status of the operational Ground Motion Service Germany, *in* 'EUSAR 2018; 12th European Conference on Synthetic Aperture Radar', Vol. 12, pp. 1–3.

Kampes, B. M. (2005), Displacement parameter estimation using permanent scatterer interferometry, Dissertation (TU Delft), Delft University of Technology.

Karakus, O., Kuruoglu, E. E. & Achim, A. (2021), A modification of Rician distribution for SAR image modelling, *in* 'EUSAR 2021; 13th European Conference on Synthetic Aperture Radar', pp. 1–6.

Krieger, G., Fiedler, H., Hajnsek, I., Eineder, M., Werner, M. & Moreira, A. (2005), TanDEM-X: Mission concept and performance analysis, *in* 'Proc. of IEEE International Geoscience and Remote Sensing Symposium', Vol. 7, pp. 4890–4893.

Lavancier, F. & Rochet, P. (2016), 'A general procedure to combine estimators', *Computational Statistics and Data Analysis* **94**, 175–192.

Lee, J., Hoppel, K., Mango, S. & Miller, A. (1993), Intensity and phase statistics of multi-look polarimetric SAR imagery, *in* 'Proc. of IEEE International Geoscience and Remote Sensing Symposium', Vol. 2, pp. 813–816.

Lege, T., Kalia, A. C. & Adam, N. (2019), Application of persistent scatterer interferometry (PSI) indicates a significant sediment sink for the Ems estuary, *in* N. Goseberg & T. Schlurmann, eds, 'Coastal Structures 2019', Vol. 2019 of *Hydraulic Engineering Repository (HENRY)*, Bundesanstalt für Wasserbau, pp. 793–802.

Lombardini, F. (2003), Differential tomography: a new framework for SAR interferometry, *in* 'Proc. of IEEE International Geoscience and Remote Sensing Symposium', Vol. 2, pp. 1206–1208.

Massonnet, D., Rossi, M., Carmona, C., Adragna, F., Peltzer, G., Feigl, K. & Rabaute, T. (1993), 'The displacement field of the landers earthquake mapped by radar interferometry', *Nature* **364**, 138–142.

Meng, Q., Ke, G., Wang, T., Chen, W., Ye, Q., Ma, Z.-M. & Liu, T.-Y. (2016), A communication-efficient parallel algorithm for decision tree, *in* D. Lee, M. Sugiyama, U. Luxburg, I. Guyon & R. Garnett, eds, 'Advances in Neural Information Processing Systems', Vol. 29, Curran Associates, Inc.

Olkin, I. & Pratt, J. W. (1958), 'Unbiased estimation of certain correlation coefficients', *The Annals of Mathematical Statistics* **29**(1), 201 – 211.

Parizzi, A. & Brcic, R. (2011), 'Adaptive InSAR stack multilooking exploiting amplitude statistics: A comparison between different techniques and practical results', *IEEE Geoscience and Remote Sensing Letters* **8**(3), 441–445.

Perissin, D. & Wang, T. (2012), 'Repeat-pass SAR interferometry with partially coherent targets', *IEEE Transactions on Geoscience and Remote Sensing* **50**(1), 271–280.

Piantanida, R. (2022), 'Sentinel-1 level 1 detailed algorithm definition', ESA-TN. SEN-TN-52-7445.

Prati, C., Rocca, F. & Guarnieri, A. (1989), Effects of speckle and additive noise on the altimetric resolution of interferometric SAR (ISAR) surveys, *in* '12th Canadian Symposium on Remote Sensing Geoscience and Remote Sensing Symposium', Vol. 4, pp. 2469–2472.

Reigber, A. & Moreira, A. (2000), 'First demonstration of airborne SAR tomography using multibaseline L-band data', *IEEE Transactions on Geoscience and Remote Sensing* **38**(5), 2142–2152.

Richman, D. (1971), 'Three dimensional, azimuth-correcting mapping radar'. patent: 4,321,601.

Rocca, F. (2007), 'Modeling interferogram stacks', *IEEE Transactions on Geoscience and Remote Sensing* **45**(10), 3289–3299.

Rosen, P., Hensley, S., Joughin, I., Li, F., Madsen, S., Rodriguez, E. & Goldstein, R. (2000), 'Synthetic aperture radar interferometry', *Proceedings of the IEEE* **88**(3), 333–382.

Schmitt, M., Ahmadi, S. A., Xu, Y., Taşkin, G., Verma, U., Sica, F. & Hänsch, R. (2023), 'There are no data like more data: Datasets for deep learning in Earth observation', *IEEE Geoscience and Remote Sensing Magazine* **11**(3), 63–97.

Seymour, M. & Cumming, I. (1994), Maximum likelihood estimation for SAR interferometry, *in* 'Proc. of IEEE International Geoscience and Remote Sensing Symposium', Vol. 4, pp. 2272–2275.

Sica, F., Gobbi, G., Rizzoli, P. & Bruzzone, L. (2021), 'ϕ-net: Deep residual learning for InSAR parameters estimation', *IEEE Transactions on Geoscience and Remote Sensing* **59**(5), 3917–3941.

Sica, F., Sanjeevamurthy, P. M. & Schmitt, M. (2023), Self-supervised learning for InSAR phase and coherence estimation, *in* 'Proc. of IEEE International Geoscience and Remote Sensing Symposium', pp. 722–725.

Touzi, R. & Lopes, A. (1996), 'Statistics of the stokes parameters and of the complex coherence parameters in one-look and multilook speckle fields', *IEEE Transactions on Geoscience and Remote Sensing* **34**(2), 519–531.

Touzi, R., Lopes, A., Bruniquel, J. & Vachon, P. (1999), 'Coherence estimation for SAR imagery', *IEEE Transactions on Geoscience and Remote Sensing* **37**(1), 135–149.

Zebker, H. A. & Goldstein, R. M. (1986), 'Topographic mapping from interferometric synthetic aperture radar observations', *J. Geophys. Res* **91**(B5), 4993–4999.

Zebker, H. & Chen, K. (2005), 'Accurate estimation of correlation in InSAR observations', *IEEE Geoscience and Remote Sensing Letters* **2**(2), 124–127.

Zebker, H. & Villasenor, J. (1992), 'Decorrelation in interferometric radar echoes', *IEEE Transactions on Geoscience and Remote Sensing* **30**(5), 950–959.

Zheng, Y., Zebker, H. & Michaelides, R. (2020), A physics-based decorrelation phase covariance model for effective decorrelation noise reduction in interferogram stacks, *in* 'Proc. of IEEE International Geoscience and Remote Sensing Symposium', pp. 16–19.

Zhu, X. X. & Bamler, R. (2010), 'Very high resolution spaceborne SAR tomography in urban environment', *IEEE Transactions on Geoscience and Remote Sensing* **48**(12), 4296–4308.

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F. & Fraundorfer, F. (2017), 'Deep Learning in remote sensing: A comprehensive review and list of resources', *IEEE Geoscience and Remote Sensing Magazine* **5**(4), 8–36.

Zink, M., Moreira, A., Bachmann, M., Rizzoli, P., Fritz, T., Hajnsek, I., Krieger, G. & Wessel, B. (2017), The global TanDEM-X DEM — a unique data set, *in* 'Proc. of IEEE International Geoscience and Remote Sensing Symposium', pp. 906–909.

A List of Abbreviations

ABW	azimuth bandwidth
ASI	Italian Space Agency
BBD	Ground Motion Service Germany
BGR	Federal Institute for Geosciences and Natural Resources
BNF	Backus–Naur form
CAESAR	Component extrAction and sElection SAR
CCG	complex, circular Gaussian
CCM	complex coherence map
CDF	cumulative distribution function
CMM	coherence magnitude map
CNN	Convolutional Neural Network
DEM	digital elevation model
D-InSAR	differential InSAR
CRLB	Cramér-Rao Lower Bound
DS	distributed scatterer
DLR	German Aerospace Center
EAP	expected *a posteriori*
EAPLSP	expected *a posteriori* less strict prior
EAPSP	expected *a posteriori* strict prior
EAPWP	expected *a posteriori* without prior
EGMS	European Ground Motion Service
ELEVATE	(Enhanced probLEm solVing with quAntum compuTErs)
ESA	European Space Agency
EVD	Eigenvalue Decomposition
FIM	Fisher information matrix

GAN Generative Adversarial Network

i.i.d. independent and identically distributed

InSAR interferometric SAR

LCMM logarithmic coherence magnitude map

LightGBM Light Gradient-Boosting Machine

LOS line of sight

MAP maximum *a posteriori*

MEDAP median *a posteriori*

ML machine learning

MLE maximum likelihood estimate

MLLSP machine learning with less strict prior

MLSP machine learning with strict prior

MLWP machine learning without prior

MSE mean squared error

NASA National Aeronautics and Space Administration

NEBN Noise Equivalent Beta Naught

NESZ Noise Equivalent Sigma Zero

NN neural network

PDF probability density function

PRF pulse-repetition frequency

PS point scatterer

PSI persistent scatterer interferometry

RBW range bandwidth

RNN Recurrent Neural Network

RMSE root mean squared error

RSF range sampling frequency

SAR Synthetic Aperture Radar

SAAR signal to azimuth ambiguity ratio

SCR signal-to-clutter ratio

SHP statistically homogenous pixels

SLC single look complex

SNR signal-to-noise ratio

SRTM Shuttle Radar Topography Mission

SVD singular value decomposition

VERITAS Venus Emissivity, Radio science, InSAR, Topography, And Spectroscopy

VISAR Venus Interferometric Synthetic Aperture Radar

XGBoost eXtreme Gradient Boosting

B About the Author

B.1 Study and Work Experience

Nico Adam received the diploma degree in electrical engineering and telecommunication science from the University of Rostock, Rostock, Germany, in 1995. Since 1995, he has been with the German Aerospace Center (DLR), Oberpfaffenhofen, Germany, where he manages InSAR projects and works as a developer and project scientist with DLR's Remote Sensing Technology Institute. Relevant projects are the Shuttle Radar Topography Mission (NASA, DLR, ASI), TerraSAR-X (DLR), Terrafirma (ESA, DLR), the Sentinel-1 Ground Motion Service Germany (BBD) (BGR, DLR), ELEVATE Quantum Computing (DLR), and VERITAS for the VISAR (NASA, DLR). In 2017, he was honored as "DLR Senior Scientist" for his application-oriented work within projects. His research interests include signal processing in the PSI framework, the use of numerical weather models to mitigate wave propagation effects in InSAR and the development of algorithms and software for advanced remote sensing radar systems.

B.2 Relevant Peer Reviewed Publications

N. Adam, 'SAR coherence estimation by composition of subsample estimates and machine learning.' in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 15957-15970, 2024.

N. Adam, 'Interferometric SAR coherence magnitude estimation by machine learning.' in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 3034-3044, 2023.

N. Adam, 'Empirical Bayesian estimation of the interferometric SAR coherence magnitude.' in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 6306-6323, 2022.

R. Reulke, N. Adam, A. Brunn, M. Cramer, C. Fischer, U. Krause, V. Spreckels, U. Weidner, C. Weyand, 'Gegenwärtige und zukünftige Normungsarbeiten des NA 005-03-02 AA 'Photogrammetrie und Fernerkundung'. in *40. Wissenschaftlich-Technische Jahrestagung der DGPF*, vol. 29, Deutsche Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation, 2020.

N. Adam, 'Methodology of a troposphere effect mitigation processor for SAR interferometry.' in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 12, pp. 5334-5344, Dec. 2019.

T. Lege, A.C. Kalia, N. Adam, 'Application of persistent scatterer interferometry (PSI) indicates a significant sediment sink for the Ems estuary.' in *Coastal Structures 2019*, pp. 793-802, Bundesanstalt für Wasserbau. Coastal Structures 2019, Hannover, Germany, 2019.

F.G. Ulmer and N. Adam, 'Characterization and improvement of the structure function estimation for application in PSI.' in *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 128, pp. 40-46, 2017, ISSN 0924-2716.

F.G. Ulmer and Nico Adam, 'A Synergy method to improve ensemble weather predictions and differential SAR interferograms.' in *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 109, pp. 98-107, 2015, ISSN 0924-2716.

K. Goel and N. Adam, 'A distributed scatterer interferometry approach for precision monitoring of known surface deformation phenomena.' in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 9, pp. 5454-5468, Sept. 2014.

K. Goel and N. Adam, 'Fusion of monostatic/bistatic InSAR stacks for urban area analysis via distributed scatterers.' in *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 4, pp. 733-737, Apr. 2014.

F. Raspini, C. Loupasakis, D. Rozos, N. Adam, S. Moretti, 'Ground subsidence phenomena in the Delta municipality region (Northern Greece): Geotechnical modeling and validation with persistent scatterer interferometry.' in *International Journal of Applied Earth Observation and Geoinformation*, vol. 28, pp. 78-89, 2014.

K. Goel and N. Adam, 'Advanced stacking of TerraSAR-X and TanDEM-X data in complex urban areas.' in *Joint Urban Remote Sensing Event 2013*, Sao Paulo, Brazil, pp. 115-118, 2013.

M. Eineder, R. Bamler, X. Cong, S. Gernhardt, T. Fritz, W.X. Zhu, U. Balss, H. Breit, N. Adam, D. Floricioiu, 'Globale Kartierung und lokale Deformationsmessungen mit den Satelliten TerraSAR-X und TanDEM-X.' in *ZFV - Zeitschrift für Geodasie, Geoinformation und Landmanagement*, pp. 75-84, Wißner-Verlag, Jan. 2013.

K. Goel and N. Adam, 'An advanced algorithm for deformation estimation in non-urban areas.' in *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 73, pp. 100-110, Elsevier, 2012.

K. Goel and N. Adam, 'High resolution deformation time series estimation for distributed scatterers using TerraSAR-X data.' in *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. I-7, pp. 29-34, Copernicus Publications, XXII ISPRS Congress, Melbourne, Australia, 2012.

K. Goel and N. Adam, 'Three dimensional positioning of point scatterers based on radargrammetry.' in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50 (6), pp. 2355-2363, IEEE - Institute of Electrical and Electronics Engineers, 2012.

S. Gernhardt, N. Adam, M. Eineder, R. Bamler, 'Potential of very high resolution SAR for persistent scatterer interferometry in urban areas.' in *Annals of GIS*, vol. 2 (16), pp. 103-111, Taylor & Francis. doi:10.1080/19475683.2010.492126, ISSN 1947-5691, 2010.

K. Goel and N. Adam, 'A Bayesian method for very high resolution multi aspect angle radargrammetry.' in *EUSAR 2010*, Aachen, Germany, 2010.

W. Liebhart, N. Adam, A. Parizzi, 'Least Squares Estimation of PSI Networks for Large Scenes with Multithreaded Singular Value Decomposition.' in *EUSAR 2010*, Aachen, Gemany, 2010.

N. Adam, A. Parizzi, M. Eineder, M. Crosetto, 'Practical persistent scatterer processing validation in the course of the Terrafirma project.' in *Journal of Applied Geophysics*, vol. 69 (1), pp 59-65, Elsevier, 2009.

S. Gernhardt, N. Adam, S. Hinz, R. Bamler, 'Appearance of Persistent Scatterers for Different TerraSAR-X Acquisition Modes.' in *Proceedings of ISPRS 2009*, pp. 1-5, ISPRS, Hannover, Germany, 2009.

M. Eineder, N. Adam, R. Bamler, N. Yague-Martinez, H. Breit, 'Spaceborne spotlight SAR interferometry with TerraSAR-X.' in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47 (5), pp. 1524-1535, IEEE - Institute of Electrical and Electronics Engineers, 2009.

R. Bamler, M. Eineder, N. Adam, X.X Zhu, S. Gernhardt, 'Interferometric potential of high resolution spaceborne SAR.' in *Photogrammetrie Fernerkundung Geoinformation*, vol. 5, pp. 403-415, E. Schweizerbartsche Verlagsbuchhandlung, 2009.

X.X. Zhu, N. Adam, R. Bamler, 'Spaceborne high resolution SAR tomography: Experiments in Urban Environment Using TerraSAR-X Data.' in *Urban Remote Sensing Joint event (JURSE)*, Shanghai, China, 2009.

X.X Zhu, N. Adam, R. Bamler, 'First demonstration of spaceborne high resolution SAR tomography in urban environment using TerraSAR-X data', in *CEOS SAR Workshop on Calibration and Validation 2008*, 2009.

N. Adam, M. Eineder, N. Yague-Martinez, R. Bamler, 'TerraSAR-X high resolution SAR-interferometry.' in *7th European Conference on Synthetic Aperture Radar*, pp. 1-4, Friedrichshafen, Germany, 2008.

T. Fritz, H. Breit, N. Adam, M. Eineder, M. Lachaise, 'Interferometric SAR processing: From TerraSAR-X to TanDEM-X.' in *7th European Conference on Synthetic Aperture Radar*, pp. 1-4, Friedrichshafen, Germany, 2008.

F.J. Meyer, S. Gernhardt, N. Adam, 'Long-term and seasonal subsidence rates in urban areas from persistent scatterer interferometry.' in *Proceedings of Urban Remote Sensing Joint Event*, Urban Remote Sensing Joint Event, Paris, France, 2007.

M. Eineder and N. Adam, 'A maximum likelihood estimator to simultaneously unwrap, geocode and fuse SAR interferograms from different viewing geometries into one digital elevation model.' in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43 (1), pp. 24-36, 2005.

M. Hubig, S. Suchandt, N. Adam, 'Equivalence of cost generators for minimum cost flow phase unwrapping.' in *J. Opt. Soc. Am. A*, vol. 19 (1), pp. 64-70, 2002.

M. Eineder, N. Adam, H. Breit, S. Suchandt, J. Holzner, B. Rabus, W. Knöpfle, 'Results from the SRTM calibration phase.' in *High Resolution Mapping from Space 2001 Proceedings*, Hannover, Germany, 2001.

M. Eineder, R. Bamler, N. Adam, H. Breit, S. Suchandt, U. Steinbrecher, 'SRTM/X-SAR interferometric processing - first results.' in *EUSAR 2000 3rd European Conference on Synthetic Aperture Radar*, pp. 233-236, VDE. EUSAR 2000, München, 2000.

S. Jonsson, N. Adam, H. Bjornsson, 'Effects of subglacial geothermal activity observed by satellite radar interferometry.' in *Geophysical Research Letters*, vol. 25, pp. 1059-1062, 1998.

R. Bamler, N. Adam, G. Davidson, D. Just, 'Noise-induced slope distortion in 2-D phase unwrapping by linear estimators with application to SAR interferometry.' in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36 (3), pp. 913-921, 1998.